Beyond Medical Expenditure:

Estimating the impact of health shocks on the welfare and socio-economic outcomes of Chinese households

Thesis to be submitted as part of the Degree of Doctor of Philosophy in Public Health at the University of Oxford
Trinity Term 2016

Candidate:
Y-Ling Chi

Supervision:
Prof. Winnie Yip
Dr. Osea Giuntella

Queen's College, Nuffield Department of Population Health
University of Oxford

Beyond Medical Expenditure:
Estimating the impact of health shocks on the welfare and socio-economic outcomes of Chinese households

Abstract

Several reviews have shown that health problems are the single most common adverse event experienced by households in many low- and middle-income countries. There is a great deal of research documenting household health expenditure and its impact on poverty. However, evidence on the costs incurred by households outside of the health care system is at best scattered.

The objective of this present study is to document the household response to illness in China using two large-scale panel household surveys and following a 'health shocks' approach. Health shocks are used to address endogeneity concerns associated with the use of alternative traditional health measures. In this manuscript, the impact of health shocks on household socio-economic status and welfare is analysed using a framework encompassing household income, medical expenditure, consumption patterns, coping strategies, and labour force supply.

We find that health shocks lead to a significant increase in medical expenditure (significant for all types of shocks), and in the case of a health shock experienced by a household head, to a decrease in income ranging around five per cent of the total household income. This decrease in income is partly explained by a statistically significant reduction in labour force supply (work hours and labour force participation) from individuals who experienced a shock (larger for women and elderly). In addition, for rural-to-urban migrant workers, health shocks are associated with a higher probability of returning. However, spouses also increase their labour force supply in response, which helps avoiding large shortfalls in income. Significant coping strategies include increase in debts and remittances; and in some cases, sale of assets. On consumption, households are mostly able to maintain consumption levels following the occurrence of a health shock (with the exception of food consumption). Nonetheless, we find a significant large negative trailing impact on consumption in subsequent years. This is in line with the literature arguing that households are risk adverse and deploy ineffective coping strategies to avoid immediate shortfalls in consumption, which generate delayed costs in the long run.

The results of the analyses carried out in this thesis highlight some of the potential channels of impoverishment due to health shocks, which might be of interest for policy makers, especially in China where a large-scale health system reform is currently taking place.

Approximately 50,000 words (excluding bibliography and appendixes, total approximately 62,000 words)

Table of Contents

List of Tables 9
List of Figures 11
List of Acronyms 12
Acknowledgements 13

Chapter 1. Introduction and Conceptual Framework 16

1.1 Background of the thesis 16
1.1.1. The impoverishing effects of ill-health 16
1.1.2. Challenges in empirical estimation and health shocks 18

1.2 Conceptual approach 20
1.2.1. Immediate costs of health shocks 22
1.2.2. Coping strategies 22
1.2.3. Short and long-term welfare impact 25

1.3. Context of research and study 26
1.3.1. Social protection in China 27
1.3.2. Previous literature on health shocks in China 28

1.4. Purpose and organisation of the thesis 30

1.5. Summary discussion 33

Chapter 2. Health shocks literature review 36

2.1. Framework and strategy for the literature selection and analysis 37
2.1.1. Purpose and scope of this review 37
2.1.2. Search Strategy 38
2.1.3. Type of information collected from included studies 41
2.1.4. Summary information of the literature 41

2.2. Defining and modelling health shocks in the literature 44
2.2.1. Construction of health shock variables 44
2.2.2. Validity of the health shock variable 47
2.2.3. Types of data 50
2.2.4. Modelling 51

2.3. Results on health shock impacts: empirical results 53
2.3.1. Results on consumption and coping strategies 54
2.3.2. Income and labour force supply responses 61
2.3.3. Other outcomes 68

2.4 Conclusions and gaps in literature 69

Chapter 3. Methods and Data 76

3.1. Establishing causal links in health studies 77
3.1.1. Selection bias or unobserved omitted variable 78
3.1.2. Other sources of endogeneity 80
3.1.3. Innovative approaches taken in this field 81
3.1.4. Health shocks 83

3.2. Methods used in the manuscript 84
3.2.1. Health shock variable 84
3.2.2. Use of panel data and fixed effects models 87
3.2.3. Are these methods good enough? 90

3.3. Research settings and data 93
3.3.1. China's rapid economic growth and health system reform 93
3.3.2 Ningxia project data 97
3.3.3. China Family Panel Studies 104

3.4. Summary of methods and challenges 108

Chapter 4. The impact of health shocks on household medical expenditure, income and coping strategies 117

4.1 Theoretical background 119
4.1.1. Loss of income 119
4.1.2. Coping strategies 120

4.2. Model and measurement 121
4.2.1. Estimation strategy 121
4.3.2. Dependent variables 122
4.2.3. Dependent variables 126

4.3. Results 130
4.3.1. Descriptive statistics 130
4.3.2. Regression estimates 133

4.4. Conclusion and discussion 139

Chapter 5. The impact of health shocks on consumption: is consumption smoothing possible? 144

5.1. Introduction 144

5.2. Data and methods 146
5.2.1. Methods 146
5.2.2. Data and health shock variable 149
5.2.3. Consumption figures 150

5.3. Results 153
5.3.1. Average impact over all four waves 153
5.3.2. Long-term impacts 155

5.4. Conclusions 158

Chapter 6. The impact of health shocks on labour force supply: a household approach 160

6.1. Background and theoretical basis 163

6.2. Methods and data 166
6.2.1. Empirical strategy 166
6.2.2. Data 170
6.2.3. Measurement 171

6.3. Results 175
6.3.1. Descriptive statistics 175
6.3.2. Regression results 180
6.3.3 Sensibility to the day threshold for health shock definition 186

6.4. Discussion and limitations 188

Chapter 7. Accounting for health shocks in the decision to return of rural to urban migrants 193

7.1. Review of potential channels of impact of health shocks on return decision 195

7.2. Methods and measurement 197
7.2.1. Methods 197
7.2.2. Data 199
7.2.3. Problem with the data in 2015 200
7.2.4. Measurement 201

7.3. Results 202
7.3.1. Descriptive information 202
7.3.2. Regression estimates 204

7.4. Conclusions and limitations 208

Chapter 8. Conclusion and discussion 213

8.1. Summary evidence of the results 214
8.1.1. What are the total immediate economic costs associated with health shocks? 214
8.1.2. What types of strategies are employed by households to cope with shocks? 216
8.1.3. What are some of the long-term consequences of shocks on households? 216
8.1.4. Contributions and comparison with existing literature 218

8.2 Policy implications 220

8.3 Limitations 223

8.4. Concluding remarks 225

Bibliography 227

Appendix A1 – Questions used from the Ningxia Project Household Survey questionnaire 238

Appendix A2 – Questions used from the CFPS (household survey questionnaire) 245

Appendix B1 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on household income 250

Appendix B2 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on health expenditure 251

Appendix B3 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on savings and assets 252

Appendix B4 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on remittances and loans 253

Appendix B5 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on total consumption 254

Appendix B6 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on NFNM consumption 255

Appendix B7 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on food consumption 256

Appendix B8 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on total consumption (using diff-in-diff strategy) 257

Appendix B9 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on health expenditure (using diff-in-diff strategy) 258

Appendix B10 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on food consumption (using diff-in-diff strategy) 259

Appendix B11 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on NFNM (using diff-in-diff strategy) 260

Appendix B12 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on own work hours using a diff-in-diff strategy 261

Appendix B13 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on own work hours using a diff-in-diff strategy *(continued)* 262

Appendix B14 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on inactivity (dummy variable) using a linear probability model 263

Appendix B15 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on inactivity (dummy variable) using a linear probability model *(continued)*..... 264

Appendix B16 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on spousal work hours using a diff-in-diff strategy..... 265

Appendix B17 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on spousal work hours using a diff-in-diff strategy *(continued)*..... 266

Appendix B18 - Full regression tables for results presented in Chapter 6 on the impact of household health shocks on work hours using a diff-in-diff strategy 267

Appendix B19 - Full regression tables for results presented in Chapter 6 on the impact of household health shocks on work hours using a diff-in-diff strategy *(continued)* 268

Appendix B20 - Full regression tables for results presented in Chapter 7 on the impact of health shocks the decision to return (own decision) (fixed effects logit)..... 269

Appendix B21 - Full regression tables for results presented in Chapter 7 on the impact of health shocks the decision to return (fixed effects logit) (own decision) *(continued)*..... 270

Appendix B22 - Full regression tables for results presented in Chapter 7 on the impact of household health shocks the decision to return (other household members) (fixed effects logit) 271

List of Tables

Table 2-1. Summary of the information collected for each article 41
Table 2-2. Types of health shocks used in the literature 44
Table 2-3. Summary of results of the literature on the full insurance theory 56
Table 2-4. Summary results on coping strategies 59
Table 2-5. Summary results on income losses 62
Table 2-6. Summary conclusions on labour force participation and supply . 65
Table 3-1. Follow-up rate for households and individuals 102
Table 3-2. Differences between drop-out and non-drop out households (Ningxia Project data) 103
Table 3-3. Differences in baseline characteristics (2010) between fully observed individuals and individuals only observed in 2010 107
Table 4-1. Summary statistics on the dependent variables used in the chapter 129
Table 4-2. Rate of shock across wealth quintiles, age groups and education attainment 131
Table 4-3. Differences in household and individual pre-shock characteristics between those exposed to shock and those not exposed 132
Table 4-4. Results of regression of wage income and health expenditure (in Yuan) 135
Table 4-5. Regression results by coping mechanism (debts and savings are in Yuan, Assets refers to the Asset Index) 137
Table 5-1. Summary statistics of the consumption variables (in Yuan, not adjusted) 152
Table 5-2. Regression results by types of consumption (in Yuan) 154
Table 5-3. Long-term effects by type of consumption (total consumption and health expenditure) 156
Table 5-4. Long-term effects by type of consumption (food consumption and NFNM consumption) 157
Table 6-1. Summary statistics for health shock and main outcomes of interest 174
Table 6-2. Rate of shock across socio-economic groups 175
Table 6-3. Comparison in individual and household baseline characteristics between individuals experiencing a health shock versus individuals not experiencing a health shock. 177
Table 6-4. Summary of estimates of the impact of health shocks on own labour force supply 181
Table 6-5. Summary of estimates of the impact of health shocks on own labour force supply (decomposition) 182
Table 6-6. Summary of estimates of the impact of health shocks on spousal labour force supply 184
Table 6-7. Summary of estimates of the impact of health shocks on spousal labour force supply (decomposition) 185
Table 6-8. Variations in estimated impact of health shocks on own work hours using different thresholds for health shock variable. 187

Table 7-1. Characteristics of return migrants versus active migrants (year before return)..... 204
Table 7-2. Return migration and own health shocks (by individual characteristics)..... 206
Table 7-3. Return migration and own health shocks (by individual characteristics) (continued)..... 206
Table 7-4. Return migration and household health shocks (by individual characteristics)..... 207

List of Figures

Figure 1-1. Conceptual framework 21
Figure 2-1. Type of outcome variables considered in the reviewed studies (total: 36 articles) 43
Figure 3-1. GNI per capita growth in China 94
Figure 3-2. Map of China locating the Ningxia Province (in red) 98
Figure 3-3. Participating provinces to CFPS (in red) 105
Figure 6-1. Number of work hours by age band, baseline (2010) 178
Figure 6-2. Participation rates by age group, baseline (2010) 178
Figure 6-3. Percentage of individuals living in households with the following number of economically active adults, baseline (2010). 179

List of Acronyms

CFPS – China Family Panel Studies
NCD – Non-Communicable Diseases
IPTW – Inverse Probability of Treatment Weighting
LMIC – Low- and Middle-Income Country
NCMS – New Cooperative Medical Scheme
NFNM – Non-Food Non-Medical (consumption)
UEBMI – Urban Employee Basic Medical Insurance
URBMI – Urban Resident Basic Medical Insurance

Acknowledgements

I am most grateful to Prof Winnie Yip and Dr Osea Giuntella for the supervision of this DPhil. This thesis has greatly benefited from your guidance, outstanding research experience and knowledge of the field. I don't think that this work would have been possible without your kind patience, especially with my ever-changing ideas and plans! It has been a great pleasure to learn under your leadership.

I would like to extend my thanks to Wei Han, Reem Hafez and Eduardo Fe for their input in this work, assistance with many aspects of data handling, Stata problem solving, and friendship. Thank you to those who have reviewed and commented on this work (including during presentations), especially to Dr. Tassot and Dr. Bisiaux. I am also excessively grateful to Dr Borowitz, Dr Cashin, Dr Bump, and Prof Flahault for their mentorship and for giving me so many opportunities to collaborate with wonderful academics and policy-makers.

I also wish to acknowledge the support of the Queen's College, the Nuffield Department of Population Health (especially Prof Sasha Shepperd) and the Blavatnik School of Government. In particular, I want to highlight the role of the Florey Scholarship and Stanley Johnson Foundation Award in funding this work. I would obviously like to express gratitude to all my friends in Oxford, Paris, La Courneuve, London and Pittsburgh for their continued support and their friendship. Specifically, I would like to thank Leila Chebbi and Maria Ironside for their enduring

encouragement and almost daily presence (and distractions!) during this DPhil. You both absolutely went above what is expected from even the best of friends.

I am most indebted to my parents who have been very positive during the course of the DPhil despite not fully grasping the nature of my work and my thesis! This also applies to my wonderful nephews, who have always managed to entertain and amuse me effortlessly. A special mention goes to my very talented siblings and their lovely partners, for their unfaltering confidence in my work and supply of smuggled cheese and saucissons that kept me well fed during harsh English winters.

Last but not least, I owe a lifelong debt to my partner (Dr!) Hector Page for his unwavering support. For the last four years, you have been an inexhaustible source of encouragement, admiration, love, laughter, magic, and silliness that has always kept my head above water.

I am infinitely grateful for this DPhil experience. When it started in 2012, I thought that I was merely preparing for another academic degree. Little did I know that I was embarking on such an adventure!

"The inherent quality of a long journey is that you always come home with something other than what you were first looking for"

- Nicolas Bouvier, Japanese Chronicles[1]

[1] Not official translation from Y-Ling Chi (from French)

Chapter 1. Introduction and Conceptual Framework

1.1 Background of the thesis

1.1.1. The impoverishing effects of ill-health

In 2000, the World Bank launched "voices of the poor" a large-scale study that gathered the views of over 60,000 people living in poverty across the world. The objective of the project was to document poverty from the perspective of the poor through a participatory approach. During interviews, participants consistently identified poor health as one of the main determinants and consequences of poverty (Narayan, Patel, Schafft, Rademacher, & Koch-Schulte, 2000; Dodd & Munck, 2001). This result is not surprising. The linkages between poverty and poor health have long been acknowledged in the literature (Bryant, 1969), and have become an undisputed concern for researchers, policy-makers and global health experts. Poverty leads to physical deprivation, poor living conditions, nutrition, and sanitation, lack of shelter, and in turn to increased vulnerability to diseases. Poverty is also an obstacle to accessing basic health care services (Dodd & Munck, 2001; Wagstaff, 2002).

Conversely, medical payments and losses of income following ill health can also drive individuals and households into poverty or push them into

deeper poverty. In many Low- and Middle-Income Countries (LMICs), health care services are still primarily financed through out-of-pocket payments, which represent a major financial burden to households (McIntyre, Thiede, Dahlgren, & Whitehead, 2006) even when strong health insurance systems are in place. For instance, Li, et al. (2014) showed that in China, the implementation of the New Medical Cooperative System in rural areas, which now covers 98.5% of the rural population, has been unsuccessful in reducing the rates of catastrophic medical payments and medical impoverishment. Besides, even if individuals decide to forego treatment (for financial reasons), households might incur important financial losses stemming from lost days of work and income.

For these reasons, illnesses are considered a major risk to individual and household welfare. Studies have shown that health problems are also the single most common adverse event experienced by households in many LMICs; ahead of climate and natural hazards, and economic and political crises (Wagstaff & Lindelow, 2014; Heltberg & Lund, 2009; Bonfrer & Gustafsson-Wright, 2016). Yet little is known about the full economic consequences of health problems from a household perspective. There are a great deal of studies documenting household expenditure on health and its impact on household poverty (*e.g.* in the context of health reforms). However, evidence on the costs associated with lost workdays, productivity and coping strategies at the household level is at best scattered (Sparrow, Van Poel, Hadiwidjaja, Yumna, Warda, & Suryahadi, 2014). McIntyre, et al.'s literature review shows that the average estimated full cost of an episode of illness ranges around 10% of total household income (McIntyre, Thiede, Dahlgren, & Whitehead, 2006). More

importantly, they emphasize that such economic studies often focus on medical expenditure, overlooking other important potential channels of costs. Some of these costs might also have persistent impacts and be responsible for a deterioration of socio-economic outcomes in the long-run (Chetty & Looney, 2006).

1.1.2. Challenges in empirical estimation and health shocks

The paucity of literature on the total economic costs associated with illnesses can be partly ascribed to the important empirical challenges associated with any analysis of the link between health and outcomes of household welfare or socio-economic status. In order to estimate the total economic costs of illnesses, researchers should be able to measure the causal impact of an episode of illness on many outcomes such as medical expenditure, income, labour force supply, consumption, debts and assets (*i.e.* the marginal increase or decrease caused solely by the occurrence of the illness). Nonetheless, as highlighted in previous studies, this estimation can be problematic due to measurement errors, simultaneity and the presence of omitted variables (Currie & Madrian, 1999). Ideally, to measure a causal impact, the health measure should be completely orthogonal to individual characteristics. This essential condition is almost always violated, and many methods have been put forward to correct for potential biases associated with this problem.

A commonly used approach in recent years promotes the use of health shocks combined with other types of techniques to control for individual characteristics (*e.g.* use of panel data, propensity score matching) (Alam &

Mahal, 2014). Using elements from existing studies on health shocks, we broadly define a health shock as the transition point from good to bad health, caused by the occurrence of a discrete, sudden and unanticipated disease or accident. Health shocks do not refer to a specific health condition and there have been many ways of constructing health shock variables in the literature. Some authors use large drops in self-reported health status or mobility as health shocks, with the assumption that a rapid and sudden decline in health status is likely to be caused by an external event (Wagstaff, 2007). Alternatively, other authors have used accidents, injuries or unanticipated conditions (such as a stroke). In general, studies using the term 'health shock' will be referring to a body of literature that intends to estimate a causal impact of health variables on socio-economic outcomes. In other words, the use of health shocks over more traditional types of health variables (e.g. self-reported illness or health status) intends to limit the endogeneity problems discussed above. By contrast to other types of corrections addressing this endogeneity concern, such as structural modelling or instrumental variables, this approach can be implemented using questions and outcomes that are more routinely collected in household surveys.

While in the last two decades, the use of health shocks has sprung up in the literature, there remains many research gaps. A large share of existing studies focuses on the impact of health shocks on consumption patterns or labour force outcomes (Alam & Mahal, 2014). Less is known on coping strategies, or the long-term impacts of health shocks. Moreover, studies often investigate the impact of health shocks on a handful of outcomes in each setting; when these outcomes are intrinsically linked and should be

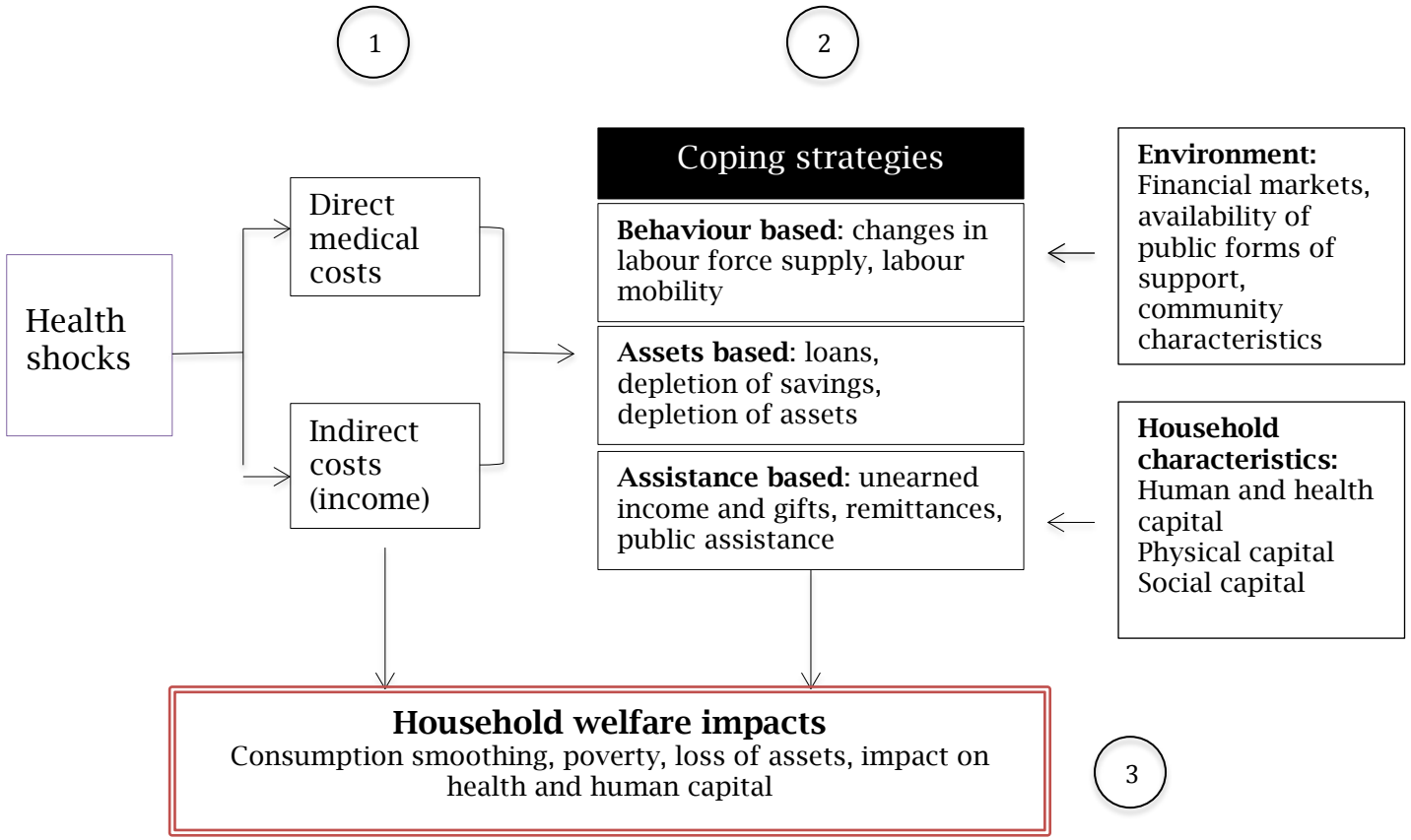
thought of as part of a broader system of household response (Mitra, Palmer, Mont, & Groce, 2016).

This manuscript intends to fill this research gap by providing an in-depth investigation of the impact of health shocks on household behaviour and welfare, employing a wide range of outcomes and recognizing their interrelatedness. In doing so, this work's final objective is to provide estimates of some of the more immediate costs incurred by illnesses, to identify household responses to cope with these costs; in order to capture different potential sources of costs associated with health shocks that are additional to medical expenditure.

1.2 Conceptual approach

The conceptual framework adopted in this manuscript is presented in Figure 1.1. This framework builds on a previous approach presented in McIntyre, Thiede, Dahlgren, & Whitehead (2006) and Sparrow, et al. (2014), with an extension to coping strategies and its determinants. Moreover, unlike McIntyre's framework, less attention is put on the health seeking behaviour following an illness.

Figure 1-1. Conceptual framework



In this manuscript, household response to health shocks is analysed in three different segments (corresponding to the numbering on Figure 1.1).

1.2.1. Immediate costs of health shocks

Firstly, the direct medical costs and loss of income are estimated in order to capture the immediate costs incurred by the occurrence of a health shock. As discussed earlier, households are likely to face important costs arising from accessing health services. These are not the sole costs associated with health shocks. The literature often separates direct health care or non-health care costs (*e.g.* medical costs, transportation costs, childkeeping costs) and indirect costs, which mostly refer to loss of income (Jo, 2014). Loss of income result from the loss of production time (including household chores and other non-wage activities) following a health shock. Depending on their 'original' contribution to households (in the absence of health shocks), this loss in production time will affect household income to varying degrees. For instance, health shocks experienced by household heads are likely to be more detrimental to household income than if experienced by an elderly parent.

1.2.2. Coping strategies

Secondly, strategies deployed by households in response to a shock (and its associated costs) will be investigated. In the development literature, coping strategies are defined as short-term choices made by households with the objective of averting the negative consequences of an event

(Sauerborn, Adams, & Hien, 1996). In the case of a health shock, shortfalls in household consumption are often identified as the main adverse consequence (Chetty & Looney, 2006). Flores, & al. (2008) emphasize the importance of the study of coping strategies in the health research literature. As argued by the authors, raw measures of catastrophic payments or medical impoverishment over-estimate the impact of medical expenditure on contemporaneous consumption.

There is no exhaustive list of coping strategies, but the most commonly cited strategies are presented in Figure 1.1. Other coping strategies might include sending children to work or taking children out of school (Dhanaraj, 2016), making low-risk choices in terms of farming (e.g. planting lower-yielding crops that are lower risk) (Chetty & Looney, 2006; Liu, 2016), or even more dramatic methods, such as the murder of an elderly parent [2] (Miguel, 2005). In this thesis, we will consider more 'traditional' types of coping strategies and mostly investigate the impact of shocks on some of the outcomes depicted above. Coping strategies can also broadly be categorised as behaviour-based, asset-based and assistance-based. This classification comes from the works of Heltberg, et al. on the anatomy of coping during crises (Heltberg, Hossain, Reva, & Turk, 2012).

There is little theoretical guidance to understand how households decide on the uptake of a strategy over another. In a study set in Burkina Faso,

[2] Miguel (2005) shows that extreme weather events in Tanzania are associated with peaks in deaths of elderly women killed by relative. The murders are partly motivated by the belief that old women ('witches') attracted the adverse climactic events upon the households. However, Miguel (2005) highlights that such murders target elderly women because they are also seen as a financial burden and liability to households.

Sauerborn, Adams, & Hien (1996) show that most households adopt at least two different types of coping strategies. As shown in Figure 1.1, the choice of coping strategies can also be influenced by many elements, including contextual elements. Existence of risk-sharing networks at the level of the community or access to financial institutions can favour the uptake of loans over other types of strategies. The existence of public assistance programs (e.g. disability or unemployment benefits, health insurance) can also play an important role in the household decision-making process.

Coping strategies can also be influenced by the type of capital available to households, and more broadly household characteristics. As shown in Figure 1-1, household's financial, human and social capital are expected to be important determinants of household response. For instance, in China, a well known phenomenon is the high private household saving rate (Chamon & Prasad, 2008; Dreger, Wang, & Zhang, 2015). Households self-insure through savings and as a result might choose to dissave in the event of a health shock. In rural areas, household members also typically work together on small-scale agricultural farms. It is likely that other household members expand their labour force supply to compensate for the loss of work days following a health shock. In the same line, another well-documented singularity of the Chinese rural economy is the high rate of temporary rural-to-urban migration. Migration decisions could therefore be influenced by a shock: sending a migrant worker can be part of a household coping strategy, or migrants might return to care for an elderly parent.

1.2.3. Short and long-term welfare impact

Finally, the last piece of analysis of the household response to health shocks relates to the short-term and more long-term impact of shocks on household welfare. In the literature, household welfare was often proxied using consumption measures. In fact, the first widespread use of the term 'health shocks' can be found in the risk and insurance literature and is concerned with the impact of health shocks on consumption patterns. Nonetheless, in recent years, there has been a recognition that consumption shortfalls are not an appropriate measure of household welfare loss following a shock. Most notably, Chetty and Looney (2006) show that households in LMICs are close to a subsistence level of consumption and that any shortfall in consumption will have a very detrimental welfare impact. Consequently, households tend to be very risk adverse and will employ any strategy (including inefficient and costly ones) to avert any reduction in levels of consumption. These coping strategies are likely to generate delayed costs or contribute to the worsening of household socio-economic outcomes in subsequent periods. Child labour or high rates of indebtness are clear exemples of inefficient coping strategies that will be detrimental to households in the long-run (Chetty & Looney, 2006). In addition, other types of delayed costs could arise from re-adjustment of household production strategy (*e.g.* favouring low-yield but low-risk crops, or labour-force transition to jobs characterised by lower-earnings). These potential long-term effects have not been investigated at all in the LMICs literature. In high income countries, labour force transitions have been shown to have persistent effects, sometimes even decades after the occurrence of the shock

(García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, 2013). These impacts are also important in explaining some of the deeper impoverishing effects of health shocks.

It is important here to note that these three segments presented above are part of a unique system to analyse the impact of health shocks. Here the word 'segments' was chosen over 'steps', as the term 'steps' introduces the idea of order or sequence of events, which is not entirely accurate in this case. One example is that households do not experience the income losses caused by health shocks all at once and subsequently decide on a set of coping strategies. On the contrary, the strategies employed by the households can influence the total costs associated with health shocks. For instance, other household members can decide to increase their labour force supply or change occupation in order to compensate for the loss of work days from the sick member. In this sense, these types of labour supply changes and occupational decisions will lead to lower loss of income (if any) and lower the overall cost of a shock.

1.3. Context of research and study

The work contained in this manuscript is set in selected provinces of China, depending on the available data. Two data sources are used in this manuscript (see Chapter 3 for full details). First of all, data from the Ningxia project (conducted in Ningxia, a relatively poor province of Western China) will be used in three of the empirical pieces. The project was launched in 2008 and aimed at evaluating changes in benefit

packages (mainly on outpatient care) of NCMS beneficiaries. An accompanying survey was conducted in 2009, 2011, 2012 and 2015 to measure the impact of the reform on roughly 22,000 individuals. In addition, the China Family Panel Survey (CFPS) is also used. The CFPS followed about 33,000 individuals in 2010 and 2012 in both rural and urban areas across 25 provinces in China (Xie & Hu, 2014).

1.3.1. Social protection in China

Since 2009, China is undertaking a systemic reform of its health care system with the goal of ensuring affordable access to basic healthcare services by 2020. Between 2009 and 2011, the Chinese government pledged an additional $125 billion dollars to expand and improve coverage of populations and health services under three different publicly funded insurance schemes (Yip, Hsiao, Chen, Hu, Ma, & Maynard, 2012). In rural areas, population coverage by the New Cooperative Medical Scheme (NCMS) is now near universal (Bai & Wu, 2014). Despite these commendable efforts, recent studies have shown that this systemic reform is failing to achieve its initial goals of financial protection (Li, et al., 2014). Sun, Jackson, Carmichael, & Sleigh (2009) have shown that out-of-pocket payment remain an important source of financial burden for households, and suggest that the structure of and low levels of reimbursement rates might be partly responsible for the observed lack of financial protection. While studying the impact of this large-scale insurance reform is not the primary goal here (and, as we will see, not possible given the timing of the data collection process), it should be kept in mind that the findings presented in this manuscript have to be

interpreted in the context of a systemic health care reform, and in most of our sample in rural areas, of near universal health care coverage.

Besides health insurance, it is worth noting that other social protection mechanisms (including unemployment or disability compensation schemes) are practically non-existent in rural areas (Juwei, 2010). In addition, as highlighted above, most rural households work on small-scale farms or migrate to find non-rural work, usually in cities and in low-skill occupations in the construction, manufacturing or service industry. These types of occupations are typically not well regulated (*e.g.* through a labour contract), although efforts to enforce more protective measures towards the employment of migrants have led to improvements in recent years (Li & Freeman, 2015).

1.3.2. Previous literature on health shocks in China

There have been three notable existing studies that overlap with some of the aspects covered in this manuscript. First of all, Lindelow & Wagstaff (2005) look at the impact of health shocks in China on household income, medical expenditure and hours of work using four waves of the China Health and Nutrition Survey (CHNS) between 1991-2000. They conclude that health shocks lead to a decrease in income by 12.4% (in the case of large health shocks) attributed to loss of work supply from the member who experiences a health shock. Another piece of work explores a similar issue and looks at changes in time use following a health shock, including impact on the spouse of the individual who experienced the shock, also

using CHNS (from 1991-2006) (Liu, 2010). This work has not been published and has been retrieved from the website of the graduate studies division of the University of California. Liu (2010) concludes that health shocks lead to important drops in work hours for men and elderly, with effects on spouse labour force supply (and important gender differentials). Finally, the very recent published study (in March 2016) of Liu (2016) should also be mentioned, which looks at the role of health insurance in the choice of coping strategy by exploiting the roll-out design for the implementation of NCMS as a natural experiment and CHNS data from 1993-2011. Liu (2016) finds that households are able to fully smooth consumption following a shock, but that health insurance helps avoid some of the more costly coping strategies, such as child labour, and decreases in investments in children and agricultural activities.

1.3.3. Contribution

With respect to this existing work, the first contribution of this thesis is that a wide range of outcomes are brought and discussed together, to depict a comprehensive picture of household response to health shocks. While Liu's work (2016) investigates the impact of coping strategies, he does not include a review of some of the more 'commonly' discussed coping strategies, such as assets, savings or loans. Liu (2010) and Lindelow & Wagstaff (2005) mostly address income and labour force status. Moreover, the data used in this thesis is different (in terms of population coverage and type of variables) and more recent (especially with regards to the work of Lindelow & Wagstaff, 2005), which is of

crucial importance given the tremendous changes in China in the recent decades. In particular, those studies were set prior to the systemic reform of the health care system, which would have an important impact on the care seeking behaviour and financial burden from medical expenditure. Most importantly, this work puts a very large emphasis on the robustness and construction of the health shock variable (using hospital admission information). Lindelow & Wagstaff (2005) and Liu (2010) use decreases in five-point scales self-reported health status. We will see that there are important concerns with regards to this measure. Similarly, Liu (2016) uses 'number of days unable to carry out work in the last 30 days' as a measure of health shock, which in itself raises concerns on the robustness and policy implication of the study. As the methods used in this manuscript are similar to those used in the studies mentioned above (except for the definition of the health shock variable), a discussion and a comparison of the results will be provided.

1.4. Purpose and organisation of the thesis

The aim of this manuscript is to document the full household consequences and response to health shocks to address the following questions:

(1) What are the total immediate economic costs associated with health shocks?
(2) What types of strategies are employed by households to cope with shocks?
(3) What are some of the long-term consequences of shocks on households?

In order to address those points, the present manuscript will be organised as follows.

Chapter 2 reviews the existing literature on health shocks in both LMICs and high-income countries. The aim of this chapter is to summarize the conclusions and findings of previous studies, but also to identify potential research gaps. Most importantly, this chapter extensively reviews the methodology used for the construction of health shock variables and the estimation techniques; and discusses the potential implication of such methodological choices for the robustness of the estimated results. In addition, this review will inform the definition of the research methods used in this manuscript.

Chapter 3 provides a short discussion of the empirical challenges associated with the estimation of such impact estimation and a justification for the use of health shocks, and describes the methods and the data used in the empirical chapters.

Chapter 4 estimates the total immediate costs of health shocks by measuring the impact of health shocks on medical expenditure and income. This chapter will also seek to identify what type of coping strategies are employed by households based on their characteristics (in particular with respect to the types of human, financial and social capital possessed by households).

In line with previous works on health shocks in the risk and insurance literature, **Chapter 5** looks at the short- and long-term impacts of health shocks (after 2, 3 and 4 years) on household consumption patterns, including on food-consumption. This chapter intends to test the 'full insurance hypothesis' that postulates that household consumption should not be responsive to health shocks. It also seeks to understand whether shocks might have a persistent negative impact on households in the long-run.

Chapter 6 will focus on labour force outcomes and review the impact of health shocks in labour supply distribution within the household. It will seek to understand the impact of shocks within the household, by documenting the changes in own labour force supply (labour force participation and work hours) following a shock, as well as potential increases or decreases in labour force supply of other household members.

One of the unique features of the Chinese rural economy (compared to other LMICs) is the high migration rate. **Chapter 7** focuses on households with migrant workers and identifies the impact of health shocks on decision to return of migrants (own health shocks and health shocks experienced by other non-migrant household members).

Finally, **Chapter 8** provides a discussion of the results presented in the empirical chapters and identify the limitations of the presented work and research gaps for future research.

1.5. Summary discussion

In this research, health shocks appear to be a frequent risk to households. From the Ningxia data, 13.2% of adults experienced at least one health shock in any of the four waves of data collection (2009-2015). This manuscript provides the first extensive review of the impact of health shocks on a comprehensive range of household welfare outcomes, covering household total estimated costs (including income losses), household consumption, labour force outcomes, and coping strategies. Some of these analyses contribute new evidence to the health shock literature (e.g. estimates of impact of health shocks on return migration, remittances, or long-term impacts).

The analysis contained in this manuscript show that health shocks lead to a substantial increase in medical expenditure and in some cases, losses of income. Borrowing seems to be the single most used coping strategy, by any socio-economic group and for any type of shock. In addition, increase in remittances and decrease in savings and sale of assets appear to be significant. Increase in remittances also appear to be a coping channel for households with a migrant worker, although the effect is only significant at the 10% level. Health shocks have a positive impact on household consumption, mostly through an increase in health expenditure. When only considering food, health shocks are associated with a significant decrease in consumption.

The observed loss of income can be attributed to the significant reduction in work hours and labour force participation following a health

shock (in particular of women, older workers and rural residents). Nonetheless, in some cases, spouses expand their labour force supply to avoid a drop of income (sometimes successfully). This points to significant private efforts within the household (from spouses) to avert shortfalls in income following a shock. Individual characteristics (age, gender, initial working status) seem to influence spouses' capacity to compensate for the loss of work time. Finally, a special analysis focusing on migrant households shows that health shocks are associated with a higher probability of return, especially for men and migrants with a severe health shock. This analysis does not support the hypothesis that migrants return to their hometowns to care for other household members in the event of a health shock in the household. Moreover, further analyses suggest that health shocks might have persistent negative effects on the long-term socio-economic status of return migrants.

Health shocks also seem to leave trailing 'delayed' costs that explain the differences in consumption several years after the shock. This is consistent with the results discussed above; in other words, households seem to adopt coping strategies to avoid shortfalls in consumption just after the shock (with the exception of a small shortfall in food consumption) that might be detrimental to households in subsequent years.

In addition, this work also highlights the differences in behaviour, response and impact on household welfare based on pre-shock individual and household characteristics, and severity of the shocks. Health shocks experienced by household heads seem to be the most detrimental to

household income and food consumption; and to lead to larger increases in borrowing. The impact of health shocks on households also seems to be very large for severe shocks that require long hospital admission.

In conclusion, the results presented in this manuscript draw attention to the economic vulnerability of Chinese households in the face of shocks. This economic vulnerability seems to arise from the uptake of potentially costly coping strategies (e.g. high uptake of debts, sale of assets, return migration), despite significant private efforts to mitigate the consequences of shocks (labour force supply transfers between household members, evidence on the use of savings). While the health insurance coverage rates in the sample were high (in the Ningxia data, only 1.5% of individuals were uninsured), health shocks also seem to lead to significant surges in health expenditure.

These results show that the consequences of health shocks are multiple, complex and need to be estimated using a wide range of outcomes. The following analyses also highlight the need to support households through targeted policies beyond the medical system. For instance, facilitating access to loans through financial institutions might be of a particular interest to policy-makers.

Chapter 2. Health shocks literature review

The first use of the term 'health shock' can be traced back to the risk and insurance literature in developing countries from beginning of the 1990s (Alderman & Paxson, 1992). From the beginning, this approach has been used to study the relationship between health and many socio-economic and health outcomes, ranging from labour force participation, smoking (Margolis, 2013), obesity (Sundmacher, 2012), to child schooling and grades (Yi, Heckman, Zhang, & Conti, 2015). Moreover, this approach is no longer exclusively applied in LMICs.

This chapter seeks to survey the rapidly expanding empirical literature on health shocks and to document the methodological approaches and conclusions put forward by previous studies. 36 articles published between 1999 and 2015 are included in this review. 14 studies (39%) were set in LMICs, while 22 (61%) were set in high-income countries. A large share of the reviewed studies were concerned with the impact of health shocks on labour force outcomes, coping strategies, and consumption patterns. The focus on the methodological aspects was used to fine-tune the approach taken in the subsequent empirical chapters and provide a benchmark against which the robustness and credibility of the results presented in this manuscript can be assessed.

This review is organised as follow. Section 2.1 provides review of the framework and strategy for the selection and analysis of the literature. Section 2.2 discusses the methodological choices for the construction of the health shock variable and modelling in the reviewed studies. Finally, Section 2.3 summarizes the conclusions put forward by the studies, and Section 2.4 outlines the main lessons from the review and the remaining literature gaps.

2.1. Framework and strategy for the literature selection and analysis

2.1.1. Purpose and scope of this review

The review carried out by Alam & Mahal (2014) is the only existing literature review of health shocks and focuses on the impoverishing effects of health shocks in LMICs. It particularly focuses on these following three measures: household out-of-pocket medical expenditure, consumption and labour force supply. 105 papers, which were published after 2000, were retrieved and included in the meta-review.

This present review is slightly different as it also covers the literature in high income countries, for which there is a rich body of evidence on the impact of health shocks on labour force outcomes. Moreover, the search and inclusion strategy for the selection of the papers is very specific in this work and pays special attention to the research methods rather than outcome measures. As a result, fewer articles are included and different dimensions are documented compared to the review of Alam & Mahal

(2014), which mostly discusses and compares results across studies. In particular, this review also provides extensive analysis of the methodological and modelling aspects of included studies. Finally, there is no strict limitation on the scope of the review and all types of socio-economic and welfare outcomes are considered.

This focus on methods is essential to assess the robustness of the conclusions in health shock studies and place this manuscript in the health shock literature. One of the central arguments in this work is that health shocks are different from other health measures *(e.g.* self-rated measure, mortality, or chronic disease) because they are constructed in order to satisfy an exogeneity assumption. This assumption is crucial to the empirical strategy as it allows researchers to consider causal impact rather than descriptive associations (full discussion in Chapter 3). If researchers consider that this assumption is violated, corrections to the model can be applied to account for the endogeneity of the health shock variable. These elements have not been documented in the work of Alam & Mahal (2014).

2.1.2. Search Strategy

The literature search was conducted using Science Direct in September 2014 and updated in February 2016. The key words 'health shocks' were used to screen titles, abstracts and keywords in the English literature. The search was limited to two categories of the literature, namely "economics, econometrics and finance" and "social science". The search in 2014

produced 205 results to which another 24 results published between September 2014 and February 2016 were added.

Article title and abstracts were then screened with the following inclusion/exclusion criteria:

- The article seeks to estimate the economic impact of health shocks using **household economic, welfare outcomes or health outcomes** (*e.g.* consumption shortfalls, loss of income, future health outcomes, child well being) or document coping strategies employed following a shock.
- The scope of the review is limited to **quantitative research**. This restriction also ensures that studies are close enough in focus and methods to be compared.
- The article should at least **mention or discuss the potential endogeneity bias**. The term 'health shock' is sometimes used quite loosely and refers to very different definitions. We allow for some level of variation in the definition to occur, but postulate that shocks are designed to address specific endogeneity problems. Therefore, authors should at least discuss the assumptions under which the variable is constructed and chosen.
- Health shocks defined using measures of chronic diseases are excluded from the analysis. For instance, Smith (1999) defines health shock as the diagnosis of a chronic disease. The use of this term seems unsuitable in that context as chronic disease are (i) typically considered endogenous to some of the key outcomes of interest (e.g. labour force participation), and (ii)

result in a gradual deterioration of a functional organ or general health status and a correspondingly gradual negative effect on health (McClellan, 1998).

- For similar reasons, studies using self-declared health shock (e.g. “has your household experienced a health shock in the last year?”) or self-reported illness in the last 14 or 30 days as a measure of health shock are not included. These measures are very broad and authors usually do not justify their methods and construction of health shock variable.
- Finally, studies relying on death of a household member as their only measure of health shock are also excluded. As highlighted in previous studies, it is likely that the household response to health shocks is different from deaths of household members, in terms of consumption and coping strategies (Grimm, 2010).

Using these criteria, the number of articles to be included in this review was narrowed down to 32 articles using an initial list of citations of 229. This selection was complemented by some studies in the literature list from the review of Alam and Mahal (2014), and by using Google Scholar. This additional step has led to the inclusion of an additional 4 articles. Working papers were not included, only papers published in peer-review journals. Annex 2-1 contains the references of the 36 studies included in this review.

2.1.3. Type of information collected from included studies

For each article, methodology of construction of health shocks, empirical strategy and conclusions of the research were documented using the variables presented in Table 2.1.

Table 2-1. Summary of the information collected for each article

Country	
If not, journal name	
Type of research question	Consumption Medical expenditure Income Labour related Health outcomes Coping strategies Other outcomes
Type of research question	
Outcome variables	
Type of health shock variable	
Prevalence of health shocks in sample	
Type of dataset, year(s) of analysis	
Empirical strategy and model specification	
Types of controls included	
Discussion of exogeneity	
Corrections for endogeneity?	Yes No
Type of corrections, if any	
Conclusions of the paper	

2.1.4. Summary information of the literature

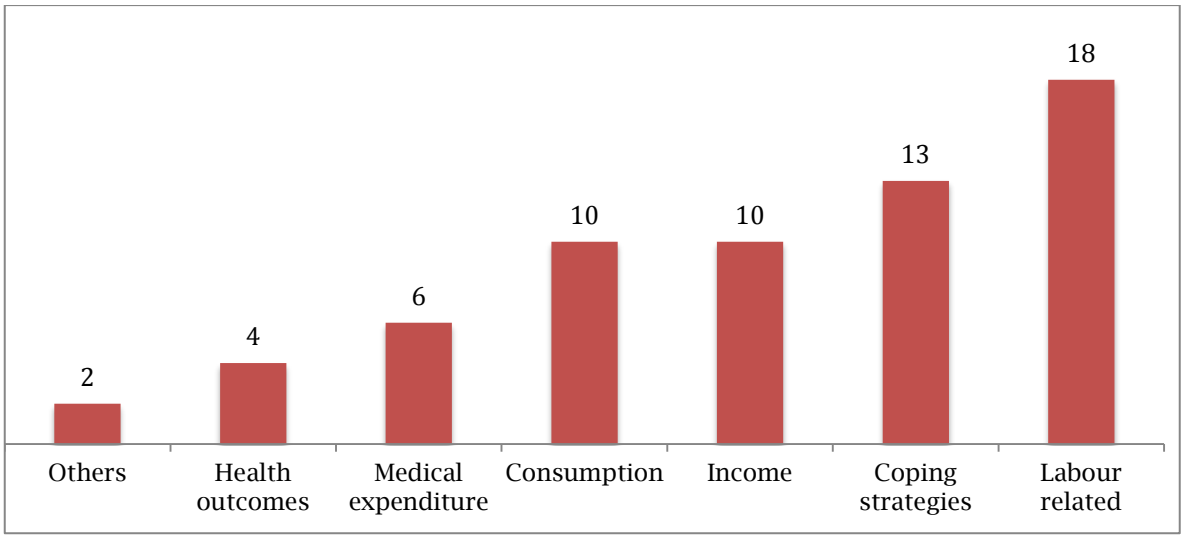
Most articles were published in health economics and policy journals or economic journals. There was no time limit on the publication; although the earliest work corresponding to this search criteria was published in 1999 (Riphahn, 1999). Nonetheless, it is worth noting that the term health

shock appeared in the literature at the beginning of the 1990s (Alderman & Paxson, 1992). These earlier works were concerned with testing the completeness of informal insurance mechanisms in communities (or the 'full insurance theory'), and therefore paid less attention to the definition and construction of the health variable. Many of these studies used self-reported health problems or death of a family member to define health shocks. In the last two decades, the use of 'health shocks' concept has expanded to many types of studies on the relationship between health and individual or household welfare, and the methods for constructing 'health shock' variables have been refined.

In our review, 22 articles (61%) were set in high-income countries (of which 6 were set in the United-States) and 14 in LMICs (38%). Most studies only looked at the impact of health shocks in one single country, although three studies (8%) sought to explain differences in response and impact (based on the country characteristics, mostly with regards to employment) in several countries (Hagan, Jones, & Rice, 2008; García-Gómez, 2011; Gupta, Kleinjans, & Larsen, 2015).

As shown on Figure 2-1, health shocks were predominantly used in studies looking at labour force outcomes (e.g. unemployment, early retirement, reduction in work hours, job turnover) (18 studies (50%)), coping strategies (36%), income and household consumption patterns (accounting for respectively 26% and 28% of studies).

Figure 2-1. Type of outcome variables considered in the reviewed studies (total: 36 articles)



Other research on non-economic outcomes include the following:

- Impact of early health shock on children outcomes (school performance and attendance) (Yi, Heckman, Zhang, & Conti, 2015).
- Impact of health shocks on other health measures or health-related behaviour (e.g. smoking) (Margolis, 2013; Sundmacher, 2012).
- Shifts in decision-making processes within the household for economic resources (change in assets portfolio) (Coile & Milligan, 2009).

2.2. Defining and modelling health shocks in the literature

2.2.1. Construction of health shock variables

Table 2-2. summarizes the type of health shock variables employed in the articles included in the review. The two most frequently used measures of health shocks are measures derived from self-rated health scores and based on the onset of specific health problems. Many studies also combined several types of health shocks in a single analysis.

Table 2-2. Types of health shocks used in the literature

Health shock variable	Number of studies	Examples
Self rated health	9	Change from good to poor health (transition)
Hospital related	8	Acute admission, admission for external causes, admissions above a day threshold
New diagnosis	8	Diagnosis of a heart condition or new cancer (except skin cancer)
Accidents or injuries	8	Road accidents, self reported severe injury
Change in mobility	8	Change in ADL score
Estimated	4	Health stock estimation (regression), deviation from age-sex cohort mean
Others	4	Medical expenditure, change in BMI, maternal complications

Nine studies (25%) define health shocks using measures of self-reported health. These indicators of self-reported health are widely available in household surveys and usually come in the form of a five-point scale question (*i.e.* "how do you rate your health in general?" with answers

ranging from very poor to very good) or a 100-point scale. In the same line, measures of decline of self-reported mobility (usually Activities of Daily Living –or ADL-) are used with the same rationale. ADLs measure individual's ability to carry out basic tasks using six dimensions of routine activities (*e.g.* bathing, dressing, walking short distances), with options for varying degrees of assistance for each task.

A health shock is usually defined by a transition from self-reported good to poor health or mobility between two rounds of survey. This method also allows researchers to define different levels of severity in health shocks and compare the results. For instance, Lindelow & Wagstaff (2005) construct two measures of health shocks using self-rated health scores (a large drop of more than two points on a five point scale versus small drop of one point). Gertler & Gruber (2002) look at shifts of consumption patterns following intermediate or severe drops in mobility.

By contrast, some studies focus directly on specific discrete health events that lead to a sudden decrease in health status or mobility. These health events are identified in the household survey (by directly asking respondents whether they have experienced a given health event) or through hospitalization data (either claims data or household survey). These health events usually include a cardio-vascular disease or accident, diagnosis of a cancer (except skin cancer), or injuries and accidents (8 studies, 20%). The definition of this variable often refers to the works of McClellan (1998) on acute hospital admissions that are considered unplanned, unreferred, relatively common and likely to lead to long-term impairments.

Finally, household surveys or medical records are sometimes used to identify health shocks defined in relation to the use of health services (most specifically hospital stay). These types of health shocks are either constructed using data on diagnosis (typically acute health events as explained above) or recorded length of stay. For instance, Wagstaff (2007) defines health shock as a hospitalisation lasting for more than three days. Defining a threshold for length of stay allows the researcher to only consider serious but non fatal health problems, which can have an important impact on both medical expenditure and other economic outcomes of household welfare (Wagstaff, 2007). In addition, several studies exclude planned hospital stays (e.g. for surgery) and only consider hospital admissions that are acute, anticipated and/or un-referred for the construction of the health shock variable.

Other measures of health shocks are also used in the reviewed literature, with less consistency. In four of the reviewed studies (11%), health shocks are estimated as deviations from the age-sex cohort in the dataset, or difference between self-declared health and estimated health stock (based on regressions using individual and household characteristics that are known to be associated with health status). All four studies use slightly different methods for constructing the estimated health shock variable.

It is also worth noting that when several measures of shocks are used in a single study, there might be important variations in estimated rates of shocks and results (magnitude of the coefficients or even patterns of

identified impact). In such studies, these differences in results are usually not discussed.

2.2.2. Validity of the health shock variable

Currie & Madrian (1999) state that defining health itself is the one of most difficult tasks faced by the researchers and that there are underpinning conceptual and methodological problems in defining any health variables. Health shocks are no exception to this rule. As shown above, there have been multiple definitions of 'health shocks' in the literature, depending on the availability of data and the intended empirical applications. This section will discuss the potential problems associated with each type of variable.

9 studies (25%) in this review use large decreases in self-reported health status as a measure of health shock. The rationale behind this choice is that large decreases in health status over a rather restricted period of time are often caused by major health events (*e.g.* an accident, injury or a stroke) and therefore more likely to be exogenous or at least unanticipated in their occurrence (Lindelow & Wagstaff, 2005). By differencing self-reported health or mobility between two waves, this method also intends to eliminate potential sources of bias stemming from unobserved time-invariant characteristics that would otherwise be uncontrolled for. Self-reported health status has been shown to be closely associated with the number of physician contacts, and a strong predictor of mortality and morbidity (Miilunpalo, Vuori, Oja, Pasanen, & Urponen, 1997; Burström & Fredlund, 2001).

Nonetheless, previous literature has suggested that the reporting of such measures can be subject to important measurement errors. In particular, authors have raised concerns about a potential 'justification bias', *i.e.* respondent's propensity to over-report health problems to justify changes in household or individual socio-economic status and labour force status (Bound, 1991). The justification bias has been particularly explored in high-income countries where employment and pension benefits can act as economic incentives to over-report illnesses (Kapteyn, Smith, & Van Soest, 2011). However, such social benefits are not widely available in the developing world. Social norms and expectations on employment could also lead to a justification bias, although no reference to support this claim was found (Baker, Stabile, & Deri, 2004).

Alternatively, using the same rationale, a large share of studies (19% of studies) also relies on measures of mobility, typically ADL. ADL is an index obtained through several questions pertaining to individual's ability to carry out several activities. ADLs are often considered a reliable outcome and show strong correlation with morbidity and hospital use (Millán-Calenti, et al., 2010). Nonetheless, they are not as widely available as self-reported health in surveys, and usually restricted to older individuals in surveys (*e.g.* adults over 45 years of age).

For these reasons, outcomes based on more precise conditions (either self-reported or from administrative data) are sometimes favoured by researchers. In addition, focusing on acute and severe conditions such as the onset of a cancer or a stroke (strategy followed in previous studies)

limits reporting errors (Jones, Rice, & Zantomio, 2015; Baker, Stabile, & Deri, 2004). Comparing self-reported and objective outcomes from hospital records and a household survey, Baker, Stabile, & Deri (2004) showed that the probability of misreporting a health condition decreases with its severity. Moreover, although correlated to individual's characteristics (including life-style, socio-economic status), the occurrence of such conditions is likely to be sudden and unpredictable (McClellan, 1998; Jones, Rice, & Zantomio, 2015). These elements suggest that with appropriate controls and corrections for other sources of bias, the use of such diagnosis can help measuring household's 'true response' to health shocks.

Some studies also use car or commuting accidents (Dano, 2005; Mohanan, 2013), and more broadly injuries, as a definition of a health shock. While the validity of these measures is not really questioned, they might be very restrictive and lack external validity (Vincent, Neilson, & Parro, 2013).

Finally, some studies define health shocks using hospital admissions, usually combined with a length of stay threshold or other types of restrictions (*e.g.* only acute or exclude surgery). There is not much justification on the definition of these thresholds and some works have used three or seven days. This method seems to be applied in the absence of information on diagnosis.

2.2.3. Types of data

With the exception of two studies, previous research on health shocks always relied exclusively or partly on a household survey. In four studies (11%) survey data is combined with administrative data sources (e.g. hospital or tax records). This strategy intends to provide more objective measures of health or income, rather than relying exclusively on self-reported data. For instance, Mohanan (2013) uses road transportation accident records to identify individuals who experienced a commuting accident and conducts two rounds of survey to estimate the changes in income and consumption.

Two studies set in the Netherlands and Austria rely exclusively on administrative data, without matching to a household survey. García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer (2013) combine national tax registries and hospital records to estimate the impact of acute hospitalizations on income and employment. Halla & Zweimüller (2013) rely solely on the Austrian social security database.

Although it was initially not a selection criteria for inclusion in the review, all the studies reviewed in this chapter make use of panel data. 13 studies (36%) use only two years and adopt a difference in difference type of specification, while 23 studies (64%) use more than three years. A large share of studies required baseline data for constructing the health

variable or matching households to control for differences in baseline characteristics.

It should be noted here that the timing of the data collection is crucial to the definition of health shocks, but this particular point has not been discussed. Some studies look at changes in health status or mobility occurring over a long period of 5 or 6 years; while others look at the same changes between two calendar years. Such differences would appear to have implications for the credibility of the constructed variable, as drops in health status over such long periods of time could be the result of different elements, some of which are not exogenous to individuals.

It is worth noting that we have come across studies using cross-sectional analysis, but they mostly relied on health shocks measures that were considered unsuitable for inclusion in this review (*e.g.* self-reported health shock). For instance, studies set in Laos and in Ethiopia rely on cross-sectional data and compare household's response and coping strategies across different types of shocks (e.g. natural disaster, economic shock and health shcoks) (Wagstaff & Lindelow, 2014; Yilma, et al., 2014).

2.2.4. Modelling

Health shocks are included as one of the exogenous regressors in almost all of the reviewed studies. The only exception is found in Genoni (2012), where prices of health inputs are used as instrumental variables for

health shocks. All other studies use fixed-effects or first difference estimations, without further corrections to account for potential endogeneity of the health shock variable. Fixed-effects or first difference models are used to control for household or individual time-invariant unobserved characteristics. In some of the studies investigating more specifically the full insurance theory, fixed effects are applied at the community level rather than at the level of households or individuals, in order to control for time invariant community unobservable characteristics.

Multiple specifications are used in studies looking at medical expenditure, debts, income or consumption figures. 18 studies use either a model in levels (with no transformation of the dependent variable) or a transformation in natural logarithms (9 studies in each category). 6 studies use a model in difference (i.e. changes in levels). Such studies tend to focus on the full-insurance theory, as it is a widely followed approach for this type of studies and was shown to be more consistent with risk-sharing theoretical models (Mace, 1991).

6 studies also combine panel data analysis with propensity score matching or other types of propensity score based techniques. Propensity scores are used to control for differences in observable baseline characteristics more thoroughly than by including covariates as controls in the regression. Matching is often applied in the context of a difference in difference type of regression (usually using two waves of data). For instance, Ambrosius & Cuecuecha (2013) investigate the impact of health shocks on loans, and match households to account for differences in

socio-economic status and characteristics of the geographical location (in terms of access to credit markets).

Four studies have also used an estimated variable of health shocks to limit measurement errors. Bound, Schoenbaum, Stinebrickner, & Waidmann (1999) provide an extensive discussion on the potential sources of bias from measurement errors. In addition, they build a method for 'purging' subjective health variables from measurement errors by regressing the subjective health variable on an extensive set of physical and mental health limitations and variables of socioeconomic status. Individual health stock is obtained by using the predicted values of this regression. Most authors employing this measure use ordered probit regression to estimate the fitted values of health stock (Hagan, Jones, & Rice, 2008; Zucchelli, Jones, Rice, & Harris, 2010). This approach is only possible in surveys where information on individual's health is detailed. Moreover it also allows researchers to combine multiple variables of individual's health into a single score, which intends to increase the accuracy of the estimates.

2.3. Results on health shock impacts: empirical results

Health shocks have been used to investigate a wide range of outcomes from household consumption, medical expenditure to labour force supply and child cognition. This section will present the results of the selected studies as follow: (i) summary of results on consumption smoothing and coping strategies, (ii) presentation of estimated impact on income and labour force outcomes, and (iii) on other outcomes.

2.3.1. Results on consumption and coping strategies

As highlighted above, the first applications of the term 'health shock' are found in the literature on risk and insurance, in particular to test the full insurance hypothesis. The full insurance hypothesis makes a very strong assumption: it predicts that informal risk-pooling arrangements will completely compensate for the costs associated with idiosyncratic risks (see Chapter 4 for full discussion). As a result, one should expect to observe no significant negative impact of health shocks on household non-medical consumption. The analysis of household behaviour following a shock should also provide evidence on risk-sharing mechanisms and existence of strong coping strategies.

This review shows very mixed results in the literature (summary findings presented in Table 2-3). Out of 10 studies, four conclude that health shocks lead to significant reduction in consumption, three show no significant impact and three provide evidence of negative impact depending on the type of consumption or households characteristics. The magnitude of impacts on household consumption can also be very large. For instance, Gertler and Gruber (2002) estimate that severe health shocks (defined as the transition from being able to perform all ADL to being able to perform none without assistance) experienced by household heads leads to a 20% reduction of non-medical consumption in Indonesia. In the same line, Asfaw and Von Braun (2004) find that non-food non-medical consumption drops by 23-28% following a health shock in Ethiopia. Other studies have found small to non-significant impact

(Powell-Jackson & Hoque, 2012; Genoni, 2012; Mitra, Palmer, Mont, & Groce, 2016). Mohanan (2013) finds significant impact on household expenditure on education, but not on other types of household consumption.

Table 2-3. Summary of results of the literature on the full insurance theory

Article reference	Impact	Summary of results
Gertler & Gruber (2002)	-	Large health shock (defined by ADLs) lowers consumption by 20%
Asfaw & Von Braun (2004)	∅ / -	No impact on food consumption, but significant reduction in non-food consumption (reduction of 23-28% of consumption growth)
Wagstaff (2007)	∅ / -	Large drop in BMI leads to reduction in food consumption in rural area (not significant in the case of hospitalization).
Gertler, Levine & Moretti (2009)	-	Severe ADL shock experienced by household heads and their wives respectively associated with a 22% and 17% reduction in household consumption (excl. medical)
Nguyet & Mangyo (2010)	-	Small but significant impact on household consumption for farm-households (not significant for non-farm households)
Genoni (2012)	∅	No significant impact on household consumption
Islam & Maitra (2012)	-	Health shocks (defined using health expenditure) have negative significant impacts on non-food consumption.
Powell-Jackson & Hoque (2012)	∅	Maternal complications reduce total household expenditure by 2.5% and food expenditure by 5%, although not significant even for severe.
Mohanan (2013)	∅ / -	Health shocks do not lead to significant reduction in consumption on food, housing, and festivals, but to small reductions in educational spending.
Mitra, Palmer, Mont & Groce (2016)	∅	Health shocks as defined by a hospitalization does not have a significant impact on any type of consumption

"-" refers to negative impact, "∅" to no impact and "∅/-" to no or negative impact (depending on group or type of consumption)

Table 2-4 presents the summary conclusions on coping strategies employed by households following a shock. The differences in results presented in the previous table (Table 2-3) could result from differences in empirical techniques. It is worth noting that studies using multiple health shock specifications in a single study also find differences in magnitude and significance of estimated coefficients. On the other hand, these mixed results could also be explained by differences in context in which studies are set, and that might influence the uptake and efficacy of coping strategies.

The full insurance hypothesis does not explicitly provide any indication of the types of risk-sharing mechanisms and coping strategies employed by households. While there is no exhaustive list of coping strategies, the most commonly cited coping strategies are savings, sale of assets, loans and debts, public transfers and gifts from friends and relatives. This review reveals that borrowing stands out as the single most often referred to coping strategy (8 studies). Debts can be contracted from formal financial institutions (Islam & Maitra, 2012; Gertler, Levine, & Moretti, 2009; Kim, Yoon, & Zurlo, 2012; Babiarz, Widdows, & Yilmazer, 2013). Other studies refer more loosely to loans, which would include borrowing from relatives and friends. This channel is likely to be important in the context of LMICs, where households typically face important borrowing constraints. Studies show moderate to very high levels of indebtedness in households. For instance, Ambrosius & Cuecuecha (2013) show that household debts double following a health shock.

Other identified coping strategies include: selling of assets (three studies), increase in transfers (including public transfers) (five studies), and use of savings (one study) (see Table 2-4). Genoni (2012) shows that transfers from non-co-resident relatives and friends increase following a shock, especially for households that experienced severe losses of income (smaller households, with severe conditions). Ambrosius & Cuecuecha (2013) compare the level of debts following a health shock between families with a migrant worker in the United-States and families without 'transnational links'. They find that the increase in levels of debt is not significant in the case of families with a migrant worker, which could be an indication that migrant members in the family send remittances in order to cover additional costs following a health shock. While it appears to be a potential coping mechanism of importance, there is not much literature documenting specifically the potential increase in remittances following a shock.

Table 2-4. Summary results on coping strategies

Article reference	Summary of results
Riphahn (1999)	Health shocks lead to increases in public transfers (*e.g.* unemployment benefits), but not sufficient to cover the losses in income
Wu (2003)	Health shocks experienced by women is associated with higher probability that households use up savings and sell assets, but for men it is insignificant.
Dano (2005)	Increase in public compensation (persistent over time: after six years: +26% for men, +37% for women)
Wagstaff (2007)	Increases in transfers and gifts from relatives and friends
Gertler, Levine & Moretti (2009)	Households that are closer to financial intuitions are better able to insure consumption, especially for households located close to a microfinance institution.
Genoni (2012)	Health shocks lead to increases in cash transfers from non-co-resident relatives and friends. Sale of "liquid assets" (e.g. jewellery or durables) do not appear to be used as a coping strategy.
Islam & Maitra (2012)	Microcredit has mitigating effects on consumption, although not very large effects. Households also appear to sell livestock following a shock.
Kim, Yoon & Zurlo (2012)	Health shocks lead to increases in unsecured debt by 6.3% ($230) to 9.3 % ($339), mainly driven by increases in medical expenditure
Powell-Jackson & Hoque (2012)	Health shocks lead to significant increase in transfers and loans
Barbiarz, Widdows & Yilmazer (2013)	Health shocks raise the probability of having unsecured debts (10% increase) and amount of unsecured debts (+11% after two years), higher impact for low-assets and uninsured households.

Ambrosius & Cuecuecha (2013)	Health shocks double household debts, but not for HH with a relative in the nuclear family working in the US - shows that remittances are an important channel to cope with medical expenditure
Mohanan (2013)	Debt was the principal mitigating mechanism households used as health shocks lead to significantly larger levels of indebtedness (difference between exposed and unexposed - 19787 rupees)
Mitra, Palmer, Mont & Groce (2016)	Loans and sale of asset seems to be the two main coping strategies (except for poor households which do not need these strategies).

"-" refers to negative impact, "Ø" to no impact and "Ø/-" to no or negative impact (depending on group or type of consumption)

2.3.2. Income and labour force supply responses

All reviewed studies unanimously showed an increase in medical expenditure, of varying degrees. Gertler & Gruber (2002) and Wagstaff (2007) show that health insurance only partly mitigates this increase in medical expenditure. In addition to these direct medical costs, households might face important losses in income. These shortfalls in income can stem from loss of labour from the household member experiencing a health shock, or from other household members (caring for the sick member). Nonetheless, other household members might increase their work supply to avoid such decreases in income. As a result, the expected impact of health shocks on income at the household level is ambiguous.

Of 10 studies looking at income, two studies find no significant effect, four studies find a significant negative impact on household income, and four studies on individual income. Studies showing a significant impact on household and individual income estimated large (and in some cases long lasting) negative effects. For instance, Gertler & Gruber (2002) find that estimated losses in income are larger than increases in medical expenditure.

These results show that health shocks have been associated with reduction in income in many settings, in LMICs and high-income countries.

Table 2-5. Summary results on income losses

Article reference	Impact	Summary of results
Gertler & Gruber (2002)	-	Reduction in earnings (magnitude not reported)
Asfaw & Von Braun (2004)	ø	No impact on income
Wu (2003)	-	Long-term effects on earnings ($2000 per year decline wage) and increased health expenditure (also of about $2000 per year).
Dano (2005)	-	Long-term effects on earnings for men and to a lesser extent for women
García-Gomez & Nicolas (2006)	-	Health shocks lead to negative impact on income (on average reduction of 1648 euros), persistent over time
Wagstaff (2007)	-	Hospitalization and reduction in BMI has a significant negative impact on household income
García-Gomez, Van Kippersluis, O'Donnell, & Van Doorslaer (2013)	-	Health shock lead to 5% loss of personal income two years after the shock, persistent over time
Genoni (2012)	-	Reduction in household income ranging from 11-18% depending on the type of shock
Powell-Jackson & Hoque (2012)	-	Reduction in household income by 2316 Takka (about $34) following a maternal complication.
Mitra, Palmer, Mont & Groce (2016)	ø	A health shock defined by a hospitalization does not have a significant impact on household income

"-" refers to negative impact, "ø" to no impact and "ø/-" to no or negative impact (depending on group or type of consumption

In this present review, health shocks have been used primarily to investigate the effects on labour force participation (50% of all reviewed studies). A vast majority of these studies are set in high-income countries, where detailed data on employment from administrative sources or household surveys is more widely available. In particular, with growing concerns on declining labour force participation, ageing populations and associated soaring burden of illness in high-income countries, research documenting the link between health shocks and labour force status has attracted the attention of many researchers and policy-makers.

All 18 studies investigating this particular question have found a significant negative impact of health shocks on individual labour force supply. The range of labour outcome considered in this review includes early retirement, hours or days of work, probability to remain in employment, exit of the labour force market. By contrast to the studies reviewed thus far, this literature highlights heterogeneity in impact depending on the type of work, age, gender, household 'position' (i.e. whether household head) and socio-economic characteristics. For instance, Bradley, Neumark, & Motika (2012) show that health shocks are associated with a decrease in the probability to remain employed, although the effects are largely mitigated by employment-contingent health insurance. They estimate that being insured increases by 19.8 percentage points the probability to remain in employment, compared to ininsured individuals who also experienced a health shock. Dano (2005) only show significant impacts for men, and Jiménez-Martín, Labeaga, &

Vilaplana Prieto (2006) only find significant effects of health shocks on early retirement for workers aged above 50.

Table 2-6. Summary conclusions on labour force participation and supply

Article reference	Impact	Summary of results
Riphahn (1999)	-	Decline in labour force participation and earnings, gender differences. Welfare state support helps mitigating negative impact for the poor but not sufficient
Gertler & Gruber (2002)	-	Severe drop in ADL would lower by 84% hours of work and increase by 74% labour force exit
Dano (2005)	-	Significant decline in employment rate in the year of the accident for men but not women
Disney, Emmerson &Wakefield (2006)	-	Estimated health shocks predict individual retirement behaviour among workers aged from 50 until state pension age.
García-Gomez &Nicolas (2006)	-	Lower by 5% the probability to remain in full time employment, and increase by 3.5% the probability to become inactive. Evidence of potential spillovers on other household members' labour force status.
Jiménez-Martín, Labeaga, & Vilaplana Prieto (2006)	-	Significant negative impact on probability to be in employment, decreases both with age and the severity of the shock, increases with educational attainment. No significant effect for under 50 years old workers.
Hagan, Jones & Rice (2008)	-	Increasing hazard of retiring, increasing with the magnitude of the health shock.
McGeary (2009)	-	Increase probability of retiring (lower effect for married individuals). Wives also respond to husband disease by increasing labour force.
Jones, Rice & Roberts (2010)	-	Health shocks associated with decision to retire early. Having an employed partner does reduce the hazard of early retirement.
Zucchelli, Jones, Rice & Harris (2010)	-	For men, health shocks increase the hazard of becoming non-employed by 50 to 320 per cent, whereas for women, by 68 to 74 per cent.

Nguyet & Mangyo (2010)	-	A large drop in ADLs experienced by household heads leads to reduction of yearly work hours by 100 and 88 hours for non-farm and farm households.
Bradley, Neumark, & Motika (2012)	-	Health shocks lead to significant reduction in probability to be in employment, but effects mitigated by employment-contingent health insurance (about a 19.8 percentage points more likely to remain in work than those without).
Powell-Jackson & Hoque (2012)	-	Maternal complications decrease work by 2-3 days per year
Cai, Mavromaras, & Oguzoglu (2013)	-	Health shocks lead to reductions in weekly work hours: on average - 2.75 hours for men and - 1.95 hours for women
García-Gomez, Van Kippersluis, O'Donnell, & Van Doorslaer (2013)	-	Health shocks lower the employment probability by seven percentage points (no recovery in later waves). Spillovers in the household: employment probability of the spouse is reduced by 1.5 percentage point
Halla & Zweimüller (2013)	-	Health shocks lead to long sickness leaves (on average 46 days), higher probability of being unemployed or to leave the labour market via disability or retirement (persistent over time) Larger impacts for women, younger individuals and for blue collar workers.
Zimmer (2013)	-	health shocks lead to 4.4 percentage points less likely to remain working full time employment in subsequent rounds and 0.7 percentage point more likely to switch to part time (differences depending on employer based insurance).
Datta Gupta, Kleinjans & Larsen (2015)	-	Health shock raises older workers' probability of not working for pay by 8.8 percentage points in Denmark and 15.9 percentage points in the United-States

"-" refers to negative impact, "Ø" to no impact and "Ø/-" to no or negative impact (depending on group or type of consumption

Three studies also investigated the impact of health shocks on other household members' labour force supply. This research question is not particularly well researched, although the impact of health shocks in this case is ambiguous. From the perspective of the literature on coping strategies, health shocks can lead to an increase in labour force supply of other household members to compensate for losses in productive time of the individual experiencing the shock. Nonetheless, household members might decrease their labour-force supply in order to provide informal care for the sick household member. Again, the observed impact is likely to be conditional on the type of work, individual characteristics, and public support. García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, (2013) estimate that health shocks decrease the probability of the spouse remaining in employment (by 1.5 percentage point); while McGeary (2009) find that wives increase their labour force supply in response to husband's illness (not significant in the case of husbands).

Finally, a few studies indicate that these negative impacts of health shocks on labour force supply are persistent in the long run. First of all, many studies show increasing effects with age, and a positive association between health shocks and decisions to retire early (Halla & Zweimüller, 2013; Hagan, Jones, & Rice, 2008; McGeary, 2009). For younger working age individuals, follow-up of individuals through long time series show persistent effects after many years (Halla & Zweimüller, 2013).

2.3.3. Other outcomes

A remaining six studies included in this review provide evidence of impacts of health shocks on other types of -mostly non-economic- outcomes. Mohanan & Maselko (2010) show that health shocks are associated with higher levels of distress (half a standard deviation higher) compared to comparison groups (constructed through matching). In addition, health shocks also appear to be associated with worsening of spousal mental health status, significant for both genders (Valle, Weeks, Taylor, & Eberstein, 2013). On the other hand, individuals experiencing a health shock are more likely to quit smoking during the same year, but with no significant impact on body mass index (Sundmacher, 2012; Margolis, 2013).

Health shocks are also associated with changes in financial portfolio composition in the United-States, as individuals who experienced a shock tend to be less risk taking and save money rather than invest in other types of assets (Coile & Milligan, 2009).

Finally, Yi, Heckman, Zhang, & Conti (2015) have shown that health shocks occuring in early childhood are associated with lower investment in child's human capital (parental response), and with long-lasting effects on child's weight, body mass index, health status and even educational attainment.

2.4 Conclusions and gaps in literature

In the last two decades, this approach using health shocks has been extended to many types of studies and outcomes. This present review provides the first discussion on the methodological choices associated with these studies (and their reliability and robustness), as well as a comprehensive review of the literature conclusions - not limited to LMICs. We have seen that while the impact of health shocks on household consumption is not clear, studies have unanimously shown a negative impact on labour force supply (including for other household members), and in most cases, negative impact on income. The magnitude of such impacts ranges from moderate to severe, and some studies have shown persistent negative impacts in the long run.

This expansion of the health shocks literature in the recent years has also led to interesting developments in terms of construction of the health shock variable and empirical strategy. For instance, the use of hospital-related information and acute medical conditions (mostly referring to the work of McClelland (1998)) have provided a solid basis for more recent studies. In addition, when used in appropriate contexts, matching individuals using propensity scores (either matching or inverse probability weighting) has helped to control for differences in socio-economic characteristics more efficiently. However, only one study in our review has explicitly tested and accounted for the endogeneity of the health shock variable (Genoni, 2012). In all other reviewed studies, the potential empirical problems associated with the use of health shocks are discussed qualitatively. Instead, these studies rely on extensive

information on the construction of the health shock variables (and its associated methodological justification) and on baseline characteristics of individuals and households experiencing a shock to support the analysis.

In addition, this review has shown many research gaps. Relatively few works investigate the health shocks in the long term, even from a more descriptive basis. Those studies that did investigate the impact of shocks on a longer time scale found long-lasting and persistent negative effects. In particular, studies looking at the full-insurance hypothesis typically only consider the immediate household response to health shocks. As pointed out by Chetty & Looney (2006), households are risk-adverse and might be able to avoid shortfalls in consumption in the short run by employing different coping strategies. Nonetheless these strategies might only 'delay' the impact of shocks on household welfare, and generate trailing costs for households. For instance, borrowing stands out as the single most used coping mechanism. Moreover, Mohanan (2013) shows that health shocks also lead to households selling livestock. This is another example of potential long-term consequences as livestock are productive assets and contribute to earnings. Longitudinal surveys in LMICs are becoming increasingly widespread and could allow for more long-term assessments of the impacts of health shocks.

Documenting coping strategies more thoroughly is also of a particular interest. So far, only a handful of health shock studies have investigated this research question, and the type of coping strategies documented is usually limited to debts, savings and selling of assets. Genoni (2012) and Ambrosius & Cuecuecha (2013) have shown that remittances might be

another potentially important coping mechanism. Nonetheless, there is almost no study investigating this particular strategy. There is also little evidence on how household characteristics or circumstances of the shocks might lead to households taking up different coping strategies.

With regards to labour force outcomes, studies set in high income countries have shown that the effects of health shocks are extremely complex and that labour force transitions following a health shock can range from early retirement, to disability, drop out of the labour force status or reduction of work hours. Moreover, studies have shown that these responses to shocks are largely dependent on household characteristics and also local characteristics of the job market and welfare support. Documenting the loss of income following a shock does not allow us to entirely grasp these changes at the individual and household level. In LMICs, understanding how shocks affect labour force decisions is therefore crucial, given the lack of social security systems and implementation of labour laws. Moreover, this analysis should be decomposed to different types of work, including rural farmers, migrant workers, urban blue collar workers etc.

Finally, most studies in LMICs are set at the household level. With regards to labour force supply response, only a handful of studies document the potential labour force response from other household members. There is also little discussion of the heterogeneity in household response based on the identity of the member who experienced a health shock. A few studies focus exclusively on shocks experienced by household heads, as they are often considered the breadwinners for the household.

Nonetheless, it would be of interest to investigate the potential differences in response for shocks experienced by other household members. This is particularly relevant to the study of the relation between health shocks and household income and labour force supply, as we should expect to observe differences in impact based on the age, education or working-status of the individual.

Annex 2-1. List of studies included in the literature review

Ambrosius, C., & Cuecuecha, A. (2013). Are remittances a substitute for credit? Carrying the financial burden of health shocks in national and transnational households. *World Development, 46*, 143-152.

Asfaw, A., & Von Braun, J. (2004). 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.

Babiarz, P., Widdows, R., & Yilmazer, T. (2013). Borrowing to cope with adverse health events: liquidity constraints, insurance coverage, and unsecured debt. *Health Economics, 22* (10), 1177-1198.

Bradley, C. J., Neumark, D., & Motika, M. (2012). The effects of health shocks on employment and health insurance: the role of employer-provided health insurance. *International Journal of Health Care Finance and Economics, 12* (4), 253-267.

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

Coile, C., & Milligan, K. (2009). How household portfolios evolve after retirement: The effect of aging and health shocks. *Review of Income and Wealth, 55* (2), 226-248.

Dano, A. M. (2005). Road injuries and long-run effects on income and employment. *Health Economics , 14* (9), 955-970.

Disney, R., Emmerson, C., & Wakefield, M. (2006). Ill health and retirement in Britain: A panel data-based analysis. *Journal of Health Economics, 25* (4), 621-649.

García Gómez, P., & Nicolás, Á. L. (2006). Health shocks, employment and income in the Spanish labour market. *Health Economics, 15* (9), 997-1009.

García-Gómez, P., Van Kippersluis, H., O'Donnell, O., & Van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources, 48* (4), 873-909.

Genoni, M. E. (2012). Health shocks and consumption smoothing: Evidence from Indonesia. *Economic Development and Cultural Change, 60* (3), 475-506.

Gertler, P., & Gruber, J. (2002). Insuring Consumption Against Illness. *The American Economic Review, 92(1), 51-70.*

Gertler, P., Levine, D. I., & Moretti, E. (2009). Do microfinance programs help families insure consumption against illness? *Health Economics, 18* (3), 257-273.

Gupta, N. D., Kleinjans, K. J., & Larsen, M. (2015). The effect of a severe health shock on work behavior: Evidence from different health care regimes. *Social Science & Medicine, 136*, 44-51.

Hagan, R., Jones, A. M., & Rice, N. (2008). Health shocks and the hazard rate of early retirement in the ECHP. *Swiss Journal of Economics and Statistics, 144* (3), 323-335.

Halla, M., & Zweimüller, M. (2013). The effect of health on earnings: Quasi-experimental evidence from commuting accidents. *Labour Economics, 24*, 23-38.

Islam, A., & 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.

Jiménez-Martín, S., Labeaga, J. M., & Prieto, C. V. (2006). A sequential model of older workers' labor force transitions after a health shock. *Health Economics, 15* (9), 1033-1054.

Jones, A. M., Rice, N., & Roberts, J. (2010). Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS. *Economic Modelling, 27* (4), 866-880.

Kim, H., Yoon, W., & Zurlo, K. A. (2012). Health Shocks, out-of-pocket medical expenses and consumer debt among middle-aged and older americans. *Journal of Consumer Affairs, 46* (3), 357-380.

Margolis, R. (2013). Health shocks in the family gender differences in smoking changes. *Journal of Aging and Health, 25(5):882-903*

McGeary, K. A. (2009). How do health shocks influence retirement decisions? *Review of Economics of the Household, 7* (3), 307-321.

Mohanan, M. (2013). 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.

Mohanan, M., & Maselko, J. (2010). Quasi-experimental evidence on the causal effects of physical health on mental health. *International Journal of Epidemiology , 39* (2), 487-493.

Nguyet, N. T., & Mangyo, E. (2010). Vulnerability of households to health shocks: an Indonesian study. *Bulletin of Indonesian Economic Studies, 46* (2), 213-235.

Mitra, S., Palmer, M., Mont, D., & Groce, N. (2016). Can households cope with health shocks in vietnam? *Health Economics, 25(7):888-907*.

Powell-Jackson, T., & Hoque, M. E. (2012). Economic consequences of maternal illness in rural Bangladesh. *Health Economics, 21* (7), 796-810.

Riphahn, R. T. (1999). Income and employment effects of health shocks A test case for the German welfare state. *Journal of Population Economics, 12* (3), 363-389.

Sundmacher, L. (2012). The effect of health shocks on smoking and obesity. *The European Journal of Health Economics, 13* (4), 451-460.

Valle, G., Weeks, J. A., & Eberstein, I. (2013). Mental and physical health consequences of spousal health shocks among older adults. *Journal of Aging and Health , 25* (7), 1121-1142.

Wagstaff, A. (2007). The economic consequences of health shocks: evidence from Vietnam. *Journal of Health Economics, 26* (1), 82-100.

Wu, S. (2003). The effects of health events on the economic status of married couples." Journal of Human Resources 38.1 (2003): 219-230. *Journal of Human Resources, 38* (1), 219-230.

YI, J., Heckman, J. J., Zhang, J., & Conti, G. (2015). Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal, 125* (588), 347-371.

Zimmer, D. M. (2015). Employment effects of health shocks: The role of fringe benefits. *Bulletin of Economic Research, 67* (4), 346-358.

Zucchelli, E., Jones, A. M., Rice, N., & Harris, A. (2010). The effects of health shocks on labour market exits: Evidence from the HILDA survey. *Australian Journal of Labour Economics, 13* (2), 191-218.

Chapter 3. Methods and Data

The development of experimental studies in social science is testament to the growing interest in identifying cause-effect relationships. Causal studies attempt to estimate the effect of changes in individual, household, or community circumstances (T) on a given outcome (Y). In other words, a causal effect is the average change in Y (ΔY) solely attributable to a unit of increase (or reduction) in T (ΔT). Policy-makers are particularly interested in causal studies, as they can help make predictions on the impact of a particular policy change on outcomes in a given population.

While this is intuitively straightforward, the estimation of a causal effect using survey data is in practice difficult (and not always possible) due to many well-documented empirical problems. Some of these, in particular the selection problem, can be addressed by random assignment of treatment in experimental studies. When random assignment is not possible (for practical or ethical reasons), researchers have to rely on survey data and build a credible and robust identification strategy. Angrist and Kruger (1999) coined the term 'identification strategy' to describe the manner with which researchers establish causal relationships between different objects using observational data and

econometric corrections to approximate as well as possible the conditions of a real experiment.

This chapter will discuss these empirical problems with an application to this manuscript, as well as their corresponding solutions. Section 3.1 provides a short overview of some of the salient questions and concerns in causal studies in health economics. Section 3.2 presents the methods used in the empirical chapters (in this manuscript) and a discussion of how well they address the potential sources of bias. A brief introduction of the research context and the two datasets used in this manuscript is provided in Section 3.3. Finally, Section 3.4 summarizes the main methodological choices and remaining concerns.

3.1. Establishing causal links in health studies

In general, the three potential problems associated with any causal study are (i) selection problem or omitted variable problem, (ii) simultaneity (or reverse causality), and (iii) measurement error. The health and socio-economic nexus is one research area where opportunities for natural experiments are quite rare, and for this reason, there have been many reviews and discussion papers focusing on these problems in this research question. For instance, the reference paper from Currie & Madrian (1999) provides an in-depth discussion and literature review on the relationship between health and labour supply.

We will start by reviewing the selection problem, as it is often considered an important threat to the validity and credibility of estimations in health shock studies.

3.1.1. Selection bias or unobserved omitted variable

For illustrative purpose, this section will look at the relationship between health problems and wage. For now, this section loosely refers to 'health problems'. Later in the chapter, it will be argued that the use of 'health shocks' over other measures of health problems intends to provide, to some extent, a solution to the selection problem.

Using a terminology consistent with the introduction, if Y_i is individual *i*'s wage and T_i individual *i*'s illness status, we would like to estimate the average change in wage (ΔY_i) caused by the change in illness status (ΔT_i) across a sample of individuals *i*=1, 2, …, I.

Ideally, we would observe the following two outcomes for a given period (t):

$$\text{Potential outcomes} = \begin{cases} Y_{0i} \; if \; T_i = 0 \\ Y_{1i} \; if \; T_i = 1 \end{cases}$$

The causal impact of having an illness for each individual would simply be the subtraction of the two outcomes (Y_{0i} - Y_{1i}).

In reality, we do not observe a given individual in two different states ($T_i = 0$ or $T_i = 1$) in the same period. The 'causal impact' is therefore usually calculated by differencing outcomes of those who have

experienced an illness and those who haven't; or to follow the terminology used in the field, the 'treated' versus 'control'. This difference contains the following two quantities: the average treatment effect on the treated, and a term referred to as the selection bias (Angrist & Pischke, 2008):

$$\underbrace{E[Y_i|T_i=1] - E[Y_i|T_i=0]}_{\text{Observed difference in average wage}} = \underbrace{E[Y_{1i}|T_i=1] - E[Y_{0i}|T_i=1]}_{\text{Average treatment effect on the treated}}$$
$$+ \underbrace{E[Y_{0i}|T_i=1] - E[Y_{0i}|T_i=0]}_{\text{Selection bias}}$$

The selection bias on T_i refers to initial differences in wages between the treated and control groups given the differences in observed and unobserved characteristics between groups. In this specific case, Grossman's (1972) seminal models of health as a human capital have illustrated some of the potential selection channels. Grossman (1972) postulates that an individual health stock requires investments in time and financial resources (*e.g.* time to exercise or financial resources to access preventive services), which are decisions jointly made with decisions with regards to labour supply and consumption. These decisions are likely to be influenced by other factors, such as preferences or risk-aversion, which cannot be observed. In this sense, there is a genuine and widely accepted theoretical basis for accounting for the influence of this selection bias when investigating the relationship between wage and health.

In a randomized control setting, T_i is assigned irrespective of individual characteristics (i.e. treatment randomization) and selection bias is therefore not a problem. Any other setting in which T_i is not randomized

might pose an issue of potential selection bias in the estimation. As Y_{0i} is not observed, it is hard to assess how large the effect of selection bias is. Angrist and Pischke (2009) suggest that this bias could be large enough to mask the real average treatment effects.

3.1.2. Other sources of endogeneity

In addition to this selection bias, two other problems are likely to arise when estimating the relationship between health and wages. First of all, researchers often run into the problem of reverse causality between health and wages. Individuals are exposed to specific health risks through their occupations, which in turn will have an impact on their own health. For instance, individuals working physically demanding jobs (*e.g.* in construction or manufacturing) might be prone to severe injuries. On the other hand, exposure to long hours of work, high levels of stress and pressure might lead to deterioration in the general health status, and vulnerability to heart disease or mental health disorders (Sparks, Cooper, Fried, & Shirom, 1997; Cooper & Marshall, 1976; Deaton, 1997; Fletcher & Sindelar, 2009).

In addition, another salient concern is the measurement error associated with the health variable. Currie & Madrian (1999) highlight that while everyone can understand the notions of good and poor health, there is no 'bulletproof' measure of health status. Age, cultural and social norms, work status and educational attainment have been shown to influence individuals' reporting of health, for both self-reported health conditions, and general health status (Butler, Burkhauser, Mitchell, & Pincus, 1987;

Kuczmarski, Kuczmarski, & Najjar, 2001; Baker, Stabile, & Deri, 2004). In addition, as discussed in Chapter 2, a problem of 'justification bias' has been identified more specifically in studies linking health and labour force outcomes. The justification bias refers to respondent's propensity to over-report health problems to justify changes in household or individual socio-economic status and labour force status (Bound, 1991). When comparing self-reported disability with more objective impairment measures, Chirikos & Nestel (1984) found evidence of differences in reporting depending on individual's work status, and concluded that the use of self-reported health in their study setting overestimated the impact on labour force supply.

3.1.3. Innovative approaches taken in this field

Despite these well-known estimation problems, the literature estimating the effect of health on socio-economic outcomes has grown considerably in the last decades, using different methods to account for the biases reviewed above.

Firstly, it should be kept in mind that the task of the researcher is complicated by the fact that only in rare circumstances is the endogeneity of a health variable testable. Testing endogeneity would require finding an appropriate instrument to health status; in our case, a variable that is correlated to health and uncorrelated to wage or other outcomes of interest. Dwyer & Mitchell (1999) use an instrumental variable approach by exploiting parental health and mortality, weight and height ratio, age, number of children and hospital admissions in the last year. Nonetheless,

few works have adopted this approach and the literature offers little insights on potential valid instruments.

Rather than using an instrumental variable approach, authors have relied on other types of empirical corrections and methods, depending on the source of endogeneity to be accounted for. Some studies have used simultaneous equation models with two correlated equations for both labour force status (or any other socio-economic outcomes) and health status (Stern, 1989; Cai & Kalb, 2006; Cai, 2010). The use of simultaneous equations is intended to reduce the bias arising from the justification bias hypothesis, reverse causality and selection effect with regards to self-reported health (Cai & Kalb, 2006). Adams, Hurd, McFadden, Merrill, & Ribeiro (2003) used Granger causality models to test the absence of causality between different health and socio-economic outcomes. Their study paved the way to many more studies published subsequently following this approach (Adda, Chandola, & Marmot, 2003). A more common approach estimates changes in socio-economic outcomes following the onset of a health problem using longitudinal surveys, sometimes combined with fixed-effects (Bender & Theodossiou, 2009). The use of fixed-effects in panel datasets can control for time-invariant individual characteristics that might influence the reporting of health and health problems (*e.g.* social norms that might influence reporting, educational attainment if looking at older adults). Other authors concerned with measurement errors focus on specific health conditions or information derived from contacts with health care services rather than general health status, sometimes relying on administrative data (*e.g.* chronic condition diagnosed by a medical doctor) rather than surveys.

3.1.4. Health shocks

In the last two decades, a new approach focused on defining health status by minimizing selection bias, and isolate as much as possible the conditions of an 'exogenous' variation in health. This approach builds on the idea of 'exogenous' variation in health capital initially presented in the seminal works of Grossman (1972) on human capital and health. Grossman showed that individual health capital is the product of initial health endowments (exogenous to individuals) and investments in health (in financial and time resources) (Grossman, 1972). Nonetheless, unlike other forms of human capital, health capital can be depleted following the occurrence of a shock.

The choice of health shocks over other types of health measures is also intended to address the problems described above. Authors either use differenced health status (to difference out time-invariant characteristics influencing the reporting of health status) or more specific health events (accidents, injuries, major acute events) to limit measurement errors (see Chapter 2 for review). Moreover, the notion of shock by definition indicates that these events are 'self-contained' or at least, discrete in time. Combined with panel data (and individual or household fixed effects) and an extensive set of controls, the use of health shocks can help to account for problems of reverse causality and better observe the timing of events (similar to a 'treatment' in experimental or quasi-experimental studies).

Finally, the construction of the shocks is crucial to address the selection problem. In general, authors argue that shocks are not strictly exogenous, but exogenous conditional upon controls (*i.e.* pre-shock household and individual controls –including matching-, fixed effects controls). Some studies also exclude unhealthy individuals at baseline (based on self-reported health or self-reporting of chronic conditions) (Pohl, Neilson, & Parro, 2014). This is done to ensure that health shocks caused by long-term conditions (that might have had an impact on socio-economic characteristics) are not included in the definition of the shock.

3.2. Methods used in the manuscript

3.2.1. Health shock variable

Main variable

In chapters 4, 5 and 7, we rely on a unique measure of health shocks, based on hospitalization data. This approach is similar to the work of Gupta, Kleinjans, & Larsenc (2015) and Jones, Rice, & Zantomio (2015) who use hospital admission information to define health shocks.

The Ningxia project data contains extensive information on individuals' health status, and their use of in-patient and out-patient care. In each wave, respondents were asked how many times they had been admitted to a hospital in the preceding 12 months. In addition, for up to two hospital admission episodes, location of the hospital, information on diagnosis, length of stay, referral, whether the person had undergone

surgery, and cost of the admission were collected for each episode of care. Each respondent is responsible for providing answers to all questions in person, and if not possible (*e.g.* if respondent is away), household heads are asked to report the information.

Respondents report the main diagnosis associated with each admission based on a list of 21 disease categories, which can be further divided in 131 more specific disease codes that are more specific. If respondents are not certain about the exact disease code, they can always just indicate the disease category rather than the code. Annex 3-1 provides the full list of disease categories and codes used in the survey. Following the works of McClelland (1998), only admissions for acute conditions are selected (*i.e.* admission for cardio-vascular disease and diagnosis of a cancer except skin cancer). In addition, the works of García-Gómez, et al. (2013) identifies two additional causes of admission that are characterized by an anticipated and sudden onset: diseases of the digestive system (*e.g.* peptic ulcers, acute gastritis and intestinal obstruction) and accidents (García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, 2013). Following their work, those two causes of admission were also included in the construction of the health shock variable. Another restriction was set: only hospitalisations that were not referred (unanticipated) were considered in the study. This definition encompasses 48% of the total hospital admission.

Among the diagnoses not included in this definition of health shocks are: infectious diseases, pregnancy, childbirth and the puerperium, respiratory diseases, mental and behavioural disorders. These diagnoses

were not included for two reasons. First, some of these conditions have been shown to be the result of a slow onset. Second, they might be related to particular attributes of individuals' socio-economic conditions as well as their living and environmental context, which we could not control for. Also, in the absence of literature on the topic, we preferred to follow a robust and well-understood approach used in other studies.

For instance, of the three main infectious diseases, Tuberculosis is by far the most common and likely to be correctly recorded in the surveyed populations. As reported by the WHO, China has an estimated one million new cases of tuberculosis [3]. Nonetheless, the living and working environment, as well as lifestyle habits, are important risk factors linked to the development of tuberculosis. Those elements are very much endogenous to many of the outcomes investigated in this manuscript, and in the absence of a sound solution to account for such endogeneity concerns, a safe option was to not include infectious diseases in the definition of the health shock variable.

The consequence of these exclusions is a rather narrow definition of health shocks, which might not be entirely representative of all hospitalizations, poor health or acute negative health events in general. However, if our estimates are biased, the results shown in this manuscript will tend to be conservative (Bound, 1991; Baker, Stabile, & Deri, 2004).

[3] See China factsheet on the website of the WHO: http://www.wpro.who.int/china/mediacentre/factsheets/tuberculosis/en

Alternative measure: length of stay

Due to the lack of information on labour supply (in particular on work hours) in the Ningxia project data, Chapter 6 uses a different dataset from other chapters and relies on the China Family Panel Studies, which also contains information on hospital admissions. While the question on primary cause of admission appears in the on-line published questionnaire, the data was not included in the final dataset shared and requests pertaining to this variable were left unanswered. As a result, information on length of stay is used to construct the health shock variable instead. In this chapter, health shocks are defined as hospital admissions lasting for seven days or more.

Other articles in the literature also use hospital admission information (Johnson & Reynolds, 2013; Wagstaff, 2007; Vincent, Neilson, & Parro, 2013). The use of a day threshold indicates that the cause of admission was severe and unanticipated (Johnson & Reynolds, 2013). It is also worth noting that the top three causes for hospitalisation in China are cardiovascular diseases, malignant tumours, lung infections (Zhang, et al., 2014), which can be considered exogenous and unanticipated (Jones, Rice, & Zantomio, 2015). The robustness of the results is also examined by running the estimations using different day thresholds ranging from 1 to 14 days (see Chapter 6).

3.2.2. Use of panel data and fixed effects models

The empirical chapters in this manuscript adopt a similar approach to the existing literature. The impact of health shocks is estimated in a fixed-effects regression or response model (if the variable of interest is binary).

Fixed-effects are either applied at the individual or at the household level, depending on the unit of analysis. The use of fixed effects is applied to control for time-invariant characteristics that are not observed.

Health shocks are introduced as exogenous variables, and the coefficients are interpreted directly from the regressions. As argued through the chapters, health shocks are used instead of traditional health variables in order to satisfy a crucial assumption in casual studies: an exogeneity assumption. While it is not argued that health shocks are random, we consider that they are exogenous conditional upon the appropriate controls and in a panel data setting (using fixed-effects).

In Chapter 6, the use of health shocks is also combined with inverse probability of treatment weighting, which are derived from propensity scores. Propensity scores were first developed by Rosenbaum & Rubin (1983) and have now become one of the most applied statistical methods in causal inference studies using observational data (with over 16 000 citations!). Propensity scores are the "probability of treatment assignment conditional on observed baseline characteristics" (Austin, 2011) and are estimated using a non-parametric equation (usually logit). There are several techniques based on propensity scores: matching, stratification, or as inverse probability of treatment weighting. Given the large sample of individuals (and relatively small number of occurrence of shocks), the propensity scores weights (included as probability weights) are used in Chapter 6. Similar techniques have been used in some instances (although not widely) in health shock studies, in order to reduce the differences in

the distribution of pre-shock characteristics between the 'treated' and the 'untreated' population (*e.g.* that experienced a shock versus no shock).

Propensity scores were considered for the other chapters 3, 4 and 7 using the Ningxia data. This idea was not applied, given that unlike the CFPS, the Ningxia data contains four waves of data. These types of corrections function best in a difference-in-difference type of study were households are evaluated before and after treatment. In the Ningxia project, the treatment is time-varying (can be received in any of the waves), which means that the propensity scores would also have to be time-varying, which is not feasible. Two solutions to this problem were available: firstly, propensity score could be calculated for households that had experienced a shock in any of the four years and kept constant across time-periods; or secondly, health shocks could be investigated in a difference-in-difference setting.

Both methods were discarded for the following reasons. While shocks are relatively 'uncommon' (on average 5.2% of the adult population for the Ningxia data in each wave, and 5.5% in the CFPS), if considering the household level and all four waves of data, they become quite frequent. Moreover, this present work seeks to analyse the effect of a shock broken down based on the identity of the individual who experienced the shock within the household (*e.g.* household head, spouse, elderly etc.), the severity and the occurrence of multiple shocks. This would have necessitated calculating a propensity score for each category of shocks and ensuring that they are correctly done.

3.2.3. Are these methods good enough?

Combining the use of fixed-effects and an extensive set of controls (and in Chapter 6, inverse probability weights) allows us to control for time-invariant unobservable and time-variant observable characteristics. The remaining source of bias that is not accounted for in this setting is time-variant unobservable characteristics; leading to omitted variables or justification biases in the model. Controlling for these omitted variables is very difficult, and in some instances impossible. In addition, in the absence of an appropriate instrumental variable, there is no procedure to test the exogeneity of the health shock variable. The evidence on the suitability of health shocks has been mostly argumentative rather than empirically tested (Alam & Mahal, 2014). Only one study in the literature review uses an instrumental variable approach to test the exogeneity of a health shock variable constructed from ADL scores (Genoni, 2012).

As shown in Chapter 2, there is no validation in the literature of one health shock measure over another and estimates and results are quite sensitive to the construction of these variables. One central lesson from the literature review is that the methodology behind the construction of the health shock variable is crucial to the validity of the causal estimation. While other measures of health status are available in our data, we choose to rely exclusively on hospital admission information for several reasons. First of all, while health shocks defined using self-assessed measures of mobility or health are widely used in the literature, they are not very specific and there is no information as to what could

have led to shocks with those measures. Consequently, it is hard to assess what the underlying assumptions are for the use of these measures. Other authors have also used 'truly' exogenous measures (*e.g.* road accidents), but the results obtained from using this variable may lack external validity due to the very narrow definition applied (Vincent, Neilson, & Parro, 2013).

It is worth noting that in this data, 25 questions are collected about the first hospital admission and 10 for the second hospital admission. This is likely to limit the influence of the justification bias, as respondents might be more likely to over-report mobility or health problems but are unlikely to fabricate a hospital admission episode altogether. Very detailed information on the location, cost, insurance reimbursement, and referral are collected. From reviewing the answers provided to the questions, there was no obvious and concerning source of inconsistency in the response.

The use of hospital admission data is also supported by literature that is not only limited to the health shock literature. The works of McClelland (1998) on hospital admission have been widely cited in this manuscript to support this methodology. In addition, the works of Baker, Stabile, & Deri (2004) also show that such large acute health events are less likely to be subject to measurement errors. Finally, another piece of work that supports this approach belongs to the literature on the health and socio-economic inequality nexus. Adams, Hurd, McFadden, Merrill, & Ribeiro (2003) investigate the causal pathways in the health and socio-economic status relationship in the United-States. They conclude that controlling

for initial health conditions, socio-economic status does not have a significant causal effect on acute conditions and accidents. On the other hand, the non-causality hypothesis could not be rejected in the case of mental health conditions, degenerative and chronic conditions (Adams, Hurd, McFadden, Merrill, & Ribeiro, 2003). It is worth noting that this work has been widely replicated (using different datasets) and questioned. Adda, Chandola, & Marmot (2003) show similar patterns in Sweden and the United-Kingdom. Nonetheless, they suggest that socio-economic conditions might have an indirect effect on acute conditions, through its impact on chronic conditions, for instance.

For Chapters 4, 5, and 7, health shocks are defined as an acute hospital admission that is considered unanticipated. These chapters also rely on a four waves of data, and taking the last wave (2015), it is therefore possible to see whether individuals who had experienced a shock were significantly different from those who had not experienced a shock. Differences in pre-shock demographic and socio-economic characteristics are discussed in individual chapters. On the other hand, several health variables are collected in the Ningxia data, and it was possible to assess whether individuals had experienced a deterioration of health using other measures of health status and mobility. Matching individuals based on age, gender, educational attainment, whether the respondent is a household head, household size and marital status, differences in health status are explored. This analysis is only available for the health shock variable used in those three chapters (Chapter 6 uses a different health shock variable and only contains two years of data). The results are presented in Annex 3-2. They show that controlling for differences in

demographic characteristics (using propensity score matching), individuals with a health shock in 2015 present significant differences in self-reported health status, eq5d, chronic disease and household head expenditure in the years preceding the shock. These differences are relatively small and sometimes insignificant when comparing 2009 health outcomes; but they gradually grow to be larger and significant in 2012 (wave preceding the shock). This might indicate that unlike the health shock literature has put forward, individuals might experience a decrease in health status in periods leading up to the health shock. This literature offers no insights on how to account for such differences; and such pre-shock health comparison is not available in any of the reviewed studies. However, it is worth noting that clinical studies have shown that such self-reported measures are significant predictors of future health events. For instance, a study from Clarke, & al. (2009) showed that EQ-5D measures were significantly associated with vascular events, complications and mortality in patients with type 2 diabetes.

This is a caveat to keep in mind while discussing the results. Such differences in health outcomes might lead to a downward bias of the estimated results.

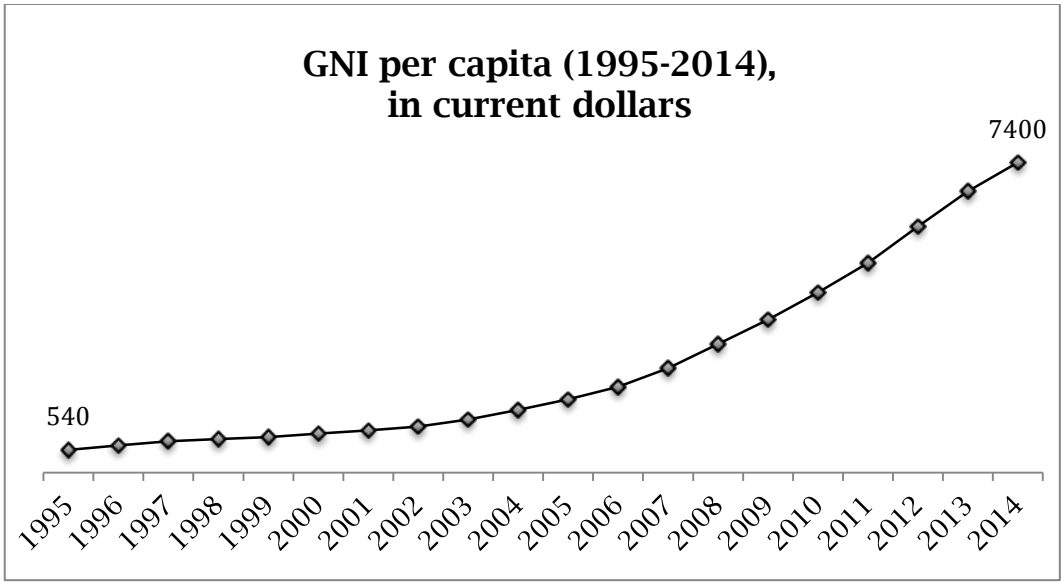
3.3. Research settings and data

3.3.1. China's rapid economic growth and health system reform

In the last three decades, the economic ascent of China has undoubtedly been one of the most notable developments in world affairs. As shown in

Figure 3-1, the GNI per capita has grown from $540 to $7400 between 1995 and 2015 (in constant dollars). Over this period, the GNI per capita has grown on average 4.6%.

Figure 3-1. GNI per capita growth in China



Source: World Bank, 2016

This remarkable economic growth has been followed by impressive gains in health outcomes nationwide, particularly in terms of infant and maternal health. Between 1995 and 2015, mortality rates for under-five fell from 45.6 to 10.4 per 1,000 (World Bank, 2016). Immunization coverage rate for Diphtheria, Pertussis and Tetanus rose from 83% to near universal (99%). Average life expectancy also grew by 6.7%. However, this general improvement in health outcomes has been uneven across the country, and major disparities in health outcomes in terms of geographic distribution have emerged during this period. In addition, chronic diseases have risen considerably as a result of both an ageing population and changes in life-styles. In 2014, the prevalence of diabetes in adults aged between 20 and 79 years old was estimated to range around 8.64%

(World Bank, 2016). In 2012, malignant neoplasms, heart diseases and cerebrovascular disease were the top three causes of deaths in China, respectively accounting for 28%, 21% and 11% of total deaths (China Centers for Disease Prevention and Control, 2012).

China is currently undergoing one of the most challenging and largest systemic reforms of its health care system. Prior to these efforts, the health care system was mostly depleted by the 1980s market-oriented economic and administrative reforms. In particular, the implementation of fiscal decentralization measures meant that local authorities received limited support for the provision of public services. The central government remained responsible for policy setting, and investments in infrastructure, but the implementation of services and funding of insurance schemes were left to provinces and counties (Uchimura & Jütting, 2009). Between 1978 and 1999, the share of central government spending in national spending fell from 32% to 15% (Blumenthal & Hsiao, 2005). These changes in funding and administrative model were particularly detrimental to the poorest regions. The dismantling of the agricultural commune system also meant that the funding of the Cooperative Medical System was withdrawn "overnight" with no replacement solution, leaving in effect 900 million rural citizens uninsured (Blumenthal & Hsiao, 2005). By 1998, 74.5% of Chinese citizens were uninsured (Chen, Liu, & Xu, 2014).

Recognizing the many adverse and extremely unfortunate consequences of these reforms on the health system, the government has dedicated monumental amounts of financial and political resources to put the

health care system back together in the last decade. In 2013, the government expenditure on health per capita was $366 (in current dollars), nearly seven times higher than in 2003, and 14 times higher than in 1996 (World Bank, 2016). Public investments also increased in relative terms: between 2003 and 2013, government expenditure on health increased from 1.75% to 3.11% of the GDP. Efforts to expand health insurance coverage were impressive. In 2013, NCMS covered 98.5% of the total rural population (Bai & Wu, 2014), and 289 million urban employees were enrolled in the Urban Employee Basic Medical Insurance (UEBMI). The urban unemployed, elderly, children and students also have the possibility to enrol in the Urban Resident Basic Medical Insurance (URBMI), the third major health insurance scheme in the country.

This research relies on data collected in selected Chinese provinces. Given the data constraints, differences in impact based on health insurance coverage are not explored in this manuscript. The first wave of data collection for the Ningxia project took place in 2008 and the coverage of our sample under NCMS was already high in this first wave. Nonetheless, it is important to bear in mind that the empirical work presented in this manuscript need to be put in the context of this health care reform. A large share of the literature investigating the impact of these reforms aims at measuring the welfare gains (including the financial protection) associated with this expansion in the health insurance coverage. This present research is somewhat complementary to the existing literature, and by identifying the long-term effects and impoverishing channels of health shocks (*e.g.* from losses of income or uptake of costly coping

strategies), it might also inform the design of better policies to complement these reforms within and outside the health care system.

It is worth noting that the work contained in this manuscript does not directly address the impact of the changes occurring in the health care system in the last decade. The first wave of data used in the manuscript was collected in 2009, and shows very high level of coverage (see discussion in 3.3.2). As a result, it is not possible for any of the chapters to test for the effects of health insurance changes on any of the outcomes presented in the manuscript. Some of those outcomes (*e.g.* saving rates, income) might be affected by the health insurance reforms, although it is unclear to which extent.

The Ningxia project data was used for all of the analyses contained in this manuscript, with the exception of those in Chapter 6. This latter relies on the China Family Panel Studies, a longitudinal survey conducted in 25 provinces (including rural and urban areas).

3.3.2 Ningxia project data

Background on the Ningxia province

The Ningxia Autonomous Province, or Ningxia province, is a small province of approximately 6.3 million inhabitants located in West China (National Bureau of Statistics of China, 2011). The population of Ningxia is by majority composed of Han Chinese, but the Hui population makes up for about a third of the total population. In the context of China, this

makes Ningxia one of the most ethnically diverse provinces in the country.

Figure 3-2. Map of China locating the Ningxia Province (in red)

Source: mapsof.net (edited to add the name of the three neighbour provinces)

The economy of the Ningxia province mainly relies on farming and grazing and is the third smallest of the country (in terms of total gross domestic product, excluding the two special administrative regions) (Ningxia Provincial Bureau of Statistics, 2013). The main crops in Ningxia are grains (wheat, rice, corn, potato, and minor cereals), oil plants, vegetable and pasture grass, most of which are still grown under a

traditional production mode, in contrast to industrial and corporate agriculture (Li, et al., N.A.).

The population of the Ningxia province is mostly rural (67.3% of the population in 2010), although the province counts 5 small cities. There are also large inequalities between urban and rural areas in the province. In 2012, average annual per capita consumption were respectively 16,674 and 5,908 Yuan in urban and rural areas (National Bureau of Statistics, 2013).

Ningxia project

The dataset used for the purpose of the analysis is a longitudinal household survey conducted in the province of Ningxia by the University of Oxford and Harvard University. This data was collected for the purpose of evaluating a health system intervention, carried out between Harvard University, the University of Oxford and the Bill and Melinda Gates Foundation, using an experimental design involving two intervention and three control counties. The intervention had the following objectives: (i) set up a single risk pool for both inpatient and outpatient services, (ii) encourage access to outpatient services by increasing reimbursement rates (including for drugs) of services, especially those sought in primary care settings (versus those in county hospitals), and (iii) replace the fee-for-service model for the payment of primary care providers to capitation method with an element of pay for performance. The intervention could affect the results if it influences the care seeking behaviour of individuals (through improved coverage) or any of the outcome variables such as savings, medical expenditure or debts.

The Ningxia survey also collects information relating to foregone hospital care (*i.e.* number of times a person was advised to be admitted to the hospital but did not). It is worth noting that this measure of foregone care does not contain any information on causes of care seeking. A year-to-year comparison between intervention and control counties shows no significant difference in rates of foregone care between intervention and control counties in any of the waves, with the exception of 2012. In 2012, of those who were advised to be hospitalised at least once, 5.8% declined admission in the control county versus 2.9% in the intervention county. There are no other significant differences in foregone hospital care for any of the other waves. As mentioned above, this measure of foregone care is very generic and does not include any information on cause of admission. We postulate that, in the event of a health shock, it is unlikely that the intervention influences the care seeking behaviour given the severity of the events and diagnoses. For instance, if an individual experiences a cardiac event, it is unlikely that he or she would only seek out-patient care or self-medicate at home. It is worth noting that there is no significant difference in the rate of recorded health shocks between the intervention and control counties in any of the waves, which further support our claim.

As this thesis defines health shocks solely using hospital information, there should be no impact of the intervention on the medical expenditure, since the intervention primarily targets access and coverage of out-patient services. Hafez (2014) finds no significant impact of the intervention beyond drug-related expenditure (explained by an increase

in access of out-patient services). Other variables, such as savings or debts might be affected to some extent by the intervention. Again, we are unable to quantitatively determine the extent of this effect. For instance, looking at debt, households in intervention and control counties are already different at baseline, with debts slightly higher in intervention counties (which are slightly richer). This trend is also reflected in post-intervention years (2011, 2012 and 2015). If these outcomes are affected by the intervention, we could argue that the effect of the intervention will only marginally impact the estimation results (*e.g.* magnitude of the estimates).

Household survey

259 villages in the five interventions and control counties were randomly selected for data collection in February of each year (2009, 2011, 2012 and 2015). Villages within counties were stratified by income and selected to be representative of county level characteristics. It is worth noting that while households were randomly selected, the sample of participants is representative of the five selected counties, rather than of the whole province.

By collecting data at the same time of the year, the survey intends to reduce seasonal variations in reporting of consumption and food production. Moreover, the month of February was chosen in order to limit non-response rates of migrant households and individuals, as traditionally, members of households are expected to return home for spring festival. This method has been used to track returned migrants in

previous surveys such as the China Health and Nutrition Survey (Popkin, Du, Zhai, & Bing, 2009).

The final sample contained 118443 observations, covering about 30,000 individuals (including roughly 21000 adults per year) in 7,000 households over the four years of measurement. As commonly encountered in longitudinal surveys, the follow-up of households is not complete across all four years (as shown in Table 3-1.)

Table 3-1. Follow-up rate for households and individuals

Follow-up	Household follow-up	Individual follow-up
All four waves	53.5%	40.52%
Three waves or more	79.76%	69.97%
Two waves or more	91.71%	86.75%
Excluding non-consecutive waves	85.33%	77.1%
Only one wave	9.29%	14.25%

In the absence of information on the reasons for attrition, a comparison in differences in means of socio-economic variables between the households observed across periods and those who dropped out is conducted. While there is no widely accepted solution for dealing with attrition problems, it can pose a problem in terms of internal validity and generalizability of the results if drop out households are selected on observable and unobservable characteristics (Alderman, Behrman, Kohler, Maluccio, & Watkins, 2001).

For each year, we divide the group between households remaining in the sample and households who dropped out between waves. A household can therefore qualify as ‘drop-out’ for the subsequent wave but be

followed up in a later year. For instance, 23% of households observed only in two waves, were observed in two non-consecutive waves. At the level of the entire sample, though, most of drop-out samples are dropped-out for all the subsequent waves.

Table 3-2. Differences between drop-out and non-drop out households (Ningxia Project data)

Baseline characteristics (year=t-1)	**Mean (drop out=0)**	**Mean (drop-out=1)**	**Difference Chi²/t-test** (1)
Age (household head, years)	47.44	44.69	0.00
Household wage income (Yuan)	15364.57	14233.65	0.00
Health expenditure (Yuan)	2236.91	2163.95	0.18
Household size	4.50	4.40	0.14
Total household consumption (Yuan)	26200.00	23483.69	0.00
Asset index	-0.75	-0.90	0.00
Migration status (household head)	29.85	41.58	0.00
No education (household head)	23.57	23.01	0.30
Farmer (household head)	71.90	67.14	0.00
Hospital shock (household head)	6.73	7.30	0.41

(1) sample size around 5,400 households (varies depending on the variable).
(2) Chi²-test performed on categorical variables, t-test on continuous variables

We observe statistically significant differences between the drop-out and followed-up households. Drop-out households are likely to be a bit younger, with lower income, household consumption, and assets. Heads of drop-out households are also more likely to be migrants and as a result less likely to be farmers. Nonetheless, there are no differences in health expenditure and hospital shocks (in the year observed before dropping out). The socio-economic differences observed between households are also likely to be related to the fact that there is a positive relationship between the age of household heads and income, assets and

consumption as households with older household heads tend to be better off than households with younger household heads.

The household survey contains detailed information on household assets and consumption, health status and health utilization of all members of the households and questions related to the labour force status and income at the household level. Household members are interviewed directly, and if they are away at the time of the interview, household heads are requested to answer on their behalf. 38.8% of questionnaires are answered directly by the individual, and 61.2% are answered by the household head. The main reason for not being directly observed is "temporary absence" (*i.e.* respondent is out of the house for work or other errands when the research team visits the house) (68.1%) and "too young" (19.4%).

3.3.3. China Family Panel Studies

In Chapter 6, we also use two waves from 2010 and 2012 from the China Family Panel Studies (CFPS), a longitudinal survey conducted by Peking University. CFPS's goal is to collect extensive information on the welfare, behaviour, lifestyles and beliefs of the Chinese population, using an integrated questionnaire for rural and urban areas. The adult questionnaire is organised around the following modules: agricultural work, marriage, cohabitation, divorce and widowhood. In addition, the family questionnaires collected information pertaining to individual characteristics (through the family roster), consumption, and work supply within the household.

Data were collected in 635 villages or urban neighbourhoods, in 162 administrative counties and 25 provinces (shown on Figure 3-3). The survey excludes the following regions/municipalities: Hong-Kong, Macao, Xinjiang, Tibet, Qinhai, Inner Mongolia, Ningxia and Hainan. Comparing the demographic characteristics of the surveyed individuals to nationwide population censuses, Xie, et al. (2014) find that the sample of surveyed individual can be considered nationally representative (excluding the regions listed above), both for urban and rural areas.

Figure 3-3. Participating provinces to CFPS (in red)

Source: mapof.net (with author's colouring of the provinces)

In 2010, CFPS covered 33,600 adults and 8,900 children who were fully interviewed out of 57,155 eligible individuals in total (response rate of 81.3%) (Xie & Hu, 2014). For eligible household members not interviewed, only basic demographic information is available. Households were visited again in 2012 and 2014 (for which data is not public at the time of writing). In 2012, full data is available for 34447 individuals (including new members to the household) and the follow-up rate was 80.6%.

Due to the use of first difference methods, only individuals observed in both waves are considered in the analysis. In this case, this drop-out might influence the results of the analysis if the cause of attrition is related to health or any of the outcome under consideration.

Again, there are significant differences in baseline socio-economic between individuals observed for two years and drop-out individuals, although it is worth noting that while significant, these differences are not always large in magnitude. Comparing the demographics between the two groups, no clear pattern emerges, so that we cannot draw a profile of the 'typical drop-out' individuals.

Table 3-3. Differences in baseline characteristics (2010) between fully observed individuals and individuals only observed in 2010

	Mean (drop out=0)	Mean (drop out=1)	t-test/ Chi2
Age	45.59	45.20	0.07
Gender (% male)	48.13	49.70	0.02
No education (%)	28.81	28.19	0.19
Han (%)	91.96	90.63	0.00
Married (%)	82.20	70.15	0.00
Residential status (%)	99.72	99.59	0.07
Schooling status (%)	6.88	5.00	0.00
Employed (%)	63.64	55.92	0.00
Urban (%)	43.31	57.69	0.00
Access to water (%)	67.23	56.23	0.22
Household (HH) size	4.26	4.14	0.00
HH consumption (Yuan)	32089	38240	0.00
HH consumption below median (%)	55.18	57.08	0.00

NB: this analysis was done on approximately 23,000 individuals

For instance, dropouts are more likely to be studying, to have no employment, to live in urban areas and to be married in 2010. Household consumption is also higher for those only observed in 2010, but a slightly higher proportion report consumption figures below the county median. Individuals missing in 2012 were also less likely to live in households with access to clean running water. This suggests that while dropout households report better outcomes overall, they tend to come from 'wealthier' counties (below county average consumption). This is consistent with the observation that the share of urban residents among dropouts is higher than for individuals observed in both years.

The survey does not contain any information as to why a given individual is not observed in 2012. Nonetheless, CFPS collects information on why individuals live away from their native household or are not considered as resident in the household. The top two reasons for living away is

studying outside of hometown and out-migration, which account for 83.9% of total reasons.

3.4. Summary of methods and challenges

This chapter intended to present with clarity the type of concerns and challenges associated with this study. In spite of the challenges discussed, the literature investigating the impact of health on socio-economic outcomes has continued to expand and has effectively become one of the most widely researched topics in the field. With very few opportunities for an experimental study, this body of literature has relied on observational data and different types of corrections to account for the problems. As a result, the analyses contained in this manuscript were carried out with the support of many studies, reviews and commentaries that have helped formulating a credible identification strategy.

The methodology for constructing the health shock variable adopted in this manuscript does not intend to generate a 'completely exogenous' shock to health. By combining this shock to the use of fixed-effects and an extensive set of controls, the intention is to provide enough 'safeguards' against the well-known biases exposed in previous works and to ensure that health shocks are exogenous conditional upon these controls. This present work also follows closely the development of the health shock literature towards more refined variables of health shocks and the use longitudinal surveys (not limited to LMICs).

There are certainly other concerns with this strategy, which will be discussed in the individual chapters as they do not necessarily only pertain to the use of health shocks, but also to the structure of the data or the specific definition of variables. In addition, some elements were not addressed in this section and will be discussed separately with each analysis. For instance, the construction of health shocks relies exclusively on information on hospitalization. There might be selection bias already arising from accessing hospital care, especially given the high cost associated with a hospital admission. There might be recall problems associated with hospital admission, especially if the questions are not answered directly by individuals. There was no solution to the problem of attrition, and imputation of data was not reasonable given the nature of the work. This could also be a source of concern for a reader. The analyses presented over the next chapters are made in full-knowledge of these limitations.

Annex 3-1. List of disease codes used in the Ningxia project data for hospital admission classification

A. Infectious Diseases	036 Hyperthyroidism
001 Typhoid and paratyphoid fevers	037 Diabetes mellitus deficiencies
002 Bacterial foodborne intoxications	039 Rickets
003 Amoebic dysentery	040 Obesity and other hyperalimentation
004 Hepatitis A	F. Diseases of the blood and bloodforming organs
005 Other intestinal infectious diseases	042 Anaemia
006 Tuberculosis	043 Other diseases of the blood and blood-forming organs
007 Tetanus	G Mental and behavioural disorders
008 Sepsis	044 Senile psychosis and pre-senile period organic psychosis
009 Measles	045 Schizophrenia
010 Japanese encephalitis	046 Depression
011 Epidemic haemorrhagic fever	047 Other mental disorders
012 Hepatitis B	H. Diseases of the nervous system
013 Leptospirosis	048 Meningitis
014 Atypical pneumonia	049 Epilepsy
015 Other parenteral infectious diseases	050 Acute infectious polyneuritis
B. Parasitic Diseases	051 Parkinson disease
016 Malaria	052 Other disorders of the nervous system
017 Schistosomiasis	I. Diseases of the eye and adnexa
018 Other parasitic diseases	053 Glaucoma
C. Malignant neoplasms	054 Cataract
019 Malignant neoplasm of nasopharynx	055 Disorders of cornea
020 Malignant neoplasm of oesophagus	056 Other disorders of eye and adnexa
021 Malignant neoplasm of stomach	J. Diseases of the ear and mastoid process
022 Malignant neoplasm of colon	057 Otitis media and mastoiditis
023 Malignant neoplasm of rectum and anus	058 Other disorders of ear and mastoid
024 Malignant neoplasm of liver	K. Diseases of the circulatory system
025 Malignant neoplasm of pancreas	059 Acute rheumatic fever
026 Malignant neoplasm of trachea, bronchus and lung	060 Chronic rheumatic heart diseases
027 Malignant neoplasm of breast	061 Angina pectoris
028 Malignant neoplasm of cervix uteri	062 Acute myocardial infarction
029 Leukaemia	063 Other ischaemic heart diseases
030 Other malignant neoplasms	064 Pulmonary heart diseases
D. In situ neoplasms, benign neoplasms and neoplasms of	065 Other forms of heart disease

uncertain or unknown behaviour	
031 Benign neoplasms of uterus	066 Hypertension
032 Benign neoplasm of brain	067 Cerebrovascular diseases
033 Other benign neoplasms	068 Varicose veins of lower extremities
034 In situ neoplasms	69 Other disorders of the circulatory system
035 Neoplasms of uncertain or unknown behaviour	L. Diseases of the respiratory system
E. Endocrine, nutritional, metabolic and immune diseases	070 Acute nasopharyngitis (common cold)
036 Hyperthyroidism	071 Acute pharyngitis, laryngitis,tonsilitis and tracheitis
072 Influenza	094 Disorders of breast
073 Pneumonia	095 Salpingitis and oophoritis
074 Chronic pharyngitis and laryngitis	096 Uterovaginal prolapse
075 Emphysema	097 Other disorders of female genital organs
76 Other chronic obstructive pulmonary diseases (COPD, chronic bronchitis)	O. Pregnancy, childbirth and the puerperium
077 Asthma	098 Spontaneous abortion
078 Other diseases of the respiratory system (acute lower respiratory infections)	099 Medical abortion
M. Diseases of the digestive system	100 Haemorrhage in pregnancy and labour
079 Diseases of oral cavity, salivary glands and jaws	101 Gestational hypertension
080 Acute and chronic gastritis	102 Normal labour
081 Peptic ulcer	103 Obstructed labour
082 Diseases of appendix	104 Complications of the puerperium
083 Abdominal hernia	105 Other diseases of pregnancy, childbirth and the puerperium
084 Intestinal obstruction	P. Diseases of the skin and subcutaneous tissue
085 Chronic hepatitis and cirrhosis	106 Carbuncle and furuncle
086 Cholelithiasis and cholecystitis	107 Dermatitis
087 Other diseases of the digestive system	108 Other disorders of the skin and subcutaneous tissue
N. Diseases of the genitourinary system	Q. Diseases of the musculoskeletal system and connective tissue
088 Nephritis and nephrosis	109 Rheumatoid arthritis
089 Pyelitis	110 Intervertebral disc disorders
090 Urolithiasis	111 Osteomyelitis
091 Other disorders of the genitourinary system	112 Other disorders of the musculoskeletal system and connective tissue
092 Hyperplasia or inflammatory diseases of prostate	R. Congenital malformations
093 Other disorders of male genital organs	113 Congenital heart diseases

094 Disorders of breast	114 Other congenital malformations
095 Salpingitis and oophoritis	S. Certain conditions originating in the perinatal period
096 Uterovaginal prolapse	115 Preterm infants
097 Other disorders of female genital organs	116 Birth trauma
	117 Fetal and birth asphyxia
O. Pregnancy, childbirth and the puerperium	118 Neonatal tetanus
088 Nephritis and nephrosis	119 Other disorders originating in the perinatal period
089 Pyelitis	T. Injury, poisoning and certain other consequences of external causes
090 Urolithiasis	120 Bone fracture
091 Other disorders of the genitourinary system	121 Dislocation, sprain and strain
092 Hyperplasia or inflammatory diseases of prostate	122 Intracranial and internal injuries (including cranial nerves)
093 Other disorders of male genital organs	118 Neonatal tetanus

119 Other disorders originating in the perinatal period	V. Factors influencing health status and contact with health services
T. Injury, poisoning and certain other consequences of external causes	127 Supervision of pregnancy
120 Bone fracture	128 Sterilization
121 Dislocation, sprain and strain	129 Persons encountering health services for specific procedures and health care
122 Intracranial and internal injuries (including cranial nerves)	130 Persons encountering health services for examination and investigation
123 Open wound and vascular injury	131 Other factors influencing health status and contact with health services
124 Burns	999 Uncertain signs, symptoms and factors
125 Poisoning and toxic effects	
126 Other sequelae of injury, poisoning	

Annex 3-2. Differences in health trajectories pre-shock (Ningxia project data)

As presented in Section 3.2, heath shocks are defined using the hospital admission data in Chapter 4, 5, and 7. The analyses contained in those chapters rely on the Ningxia project data, which contain four waves of data collected in 2009, 2011 and 2015.

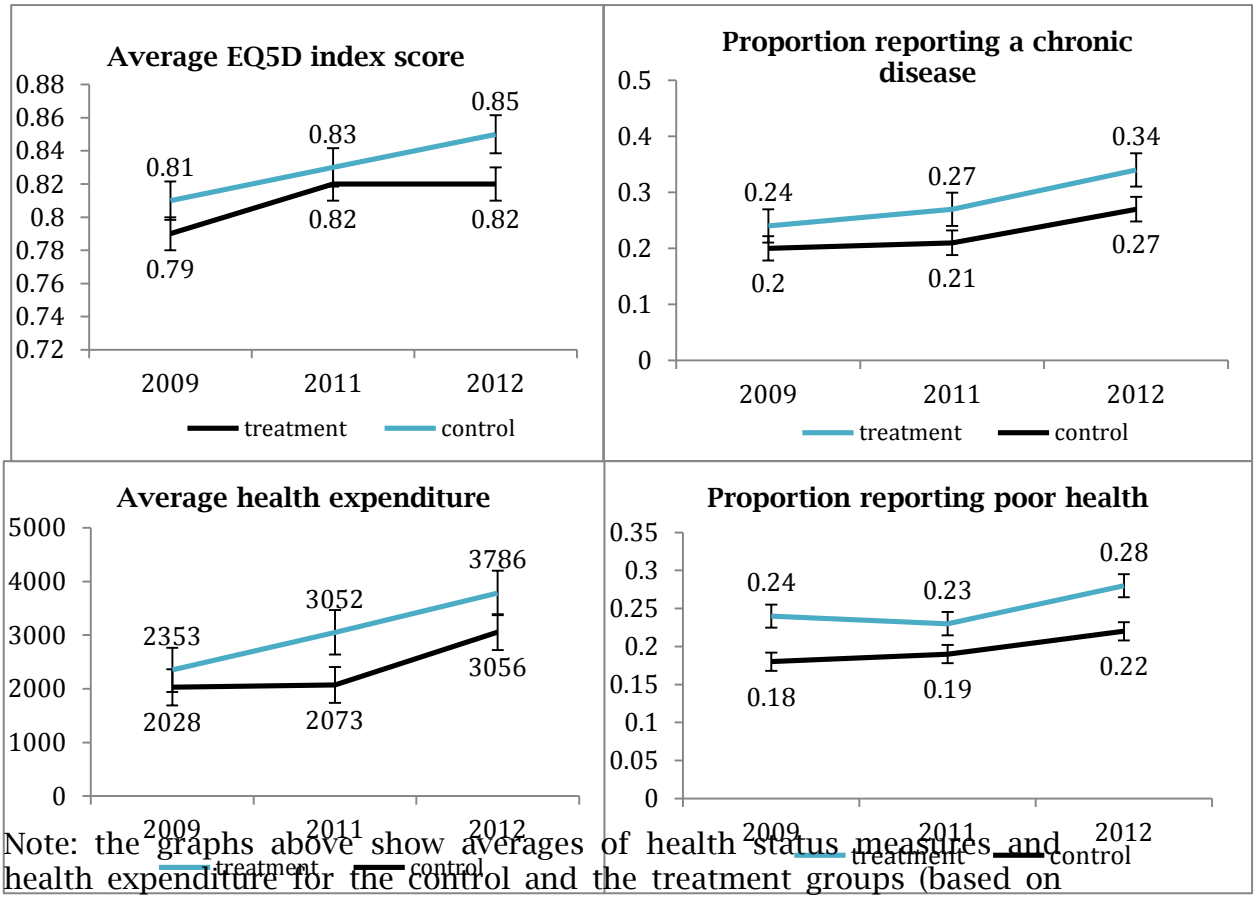
For this variable only, it is possible to comment on the pre-shock health trajectories of individuals with a health shock versus those without a health shock. In order to do so, 2015 is taken as the reference year, and for that year, a treatment and control group are defined based on whether the individual has experienced a health shock in 2015 or not. No restrictions are made on health shocks experienced in any of the previous waves: both individuals in the treatment and control group might have experienced a shock in 2009, 2011 or 2012.

Three other variables of individual health status are available in the survey.

1. Individuals are asked to rate their health as 'very good', 'good', 'fair', 'poor' and 'very poor'. From this variable, a dummy variable for 'self-reported poor health' is constructed and equals to one if individuals reported poor or very poor health (and zero otherwise).
2. Individuals are also asked to report any chronic disease formally diagnosed by a doctor (dummy variable).
3. Finally, an eq5d module collects information on general health status of individuals using five questions (see Appendix 1 for full

description of the questions). An eq5d index was computed using the eq5d command developed by Ramos-Goni & Rivero-Arias (2011). The Index is equal to 1 if individuals have a full health. It is worth noting that the questionnaire was only administered if the individual was present at the time of the interview, so it suffers from 41% non-response.

Individuals in the treatment are matched to three individuals in the control group using propensity score matching, based on educational attainment, marital status, gender, age, whether the individual is a household head and household size (those characteristics are significantly different between the treatment and control group, see Chapter 4 for full discussion). The calipher (*i.e.* maximum 'distance' or difference allowed between individuals) is set at 0.2 standard deviations, which is the usual practice in the literature using propensity score matching (Lunt, 2014). Adjusted comparisons in means (using propensity score matching) are presented in the graphs below for the three health variables and health expenditure, and in all waves preceding 2015.



Note: the graphs above show averages of health status measures and health expenditure for the control and the treatment groups (based on whether or not they have experienced a health shock in the 2015 wave) for previous waves to the shock. The groups are created using propensity score matching (see above). Confidence intervals are added for each data point.

As shown on the graph above, there are significant differences between treated and control individuals in health status in pre-shock years. These differences seem to be smaller and sometimes insignificant (in the case of Eq5d and health expenditure) in 2009 (six years prior to the shock). Nonetheless, these differences grow gradually across years and are the largest in the year prior to the health shock. For instance, in 2012, the proportion of individuals reporting poor health is six percentage points higher (or 23% larger) for individuals who have experienced a shock in

2015 compared to those who did not experience a shock. The literature offers little insights on how to account for such differences and/or whether they do have an important impact on the estimated results.

Chapter 4. The impact of health shocks on household medical expenditure, income and coping strategies

In the last two decades, both the empirical and theoretical literature has sought to document how households and individuals change their behaviour following a shock, and most specifically in the event of health shock. As shown in previous studies, health shocks are amongst the most unpredictable and sizeable shocks to household income and welfare (Wagstaff & Lindelow, 2014; Dercon & Krishnan, 2000; Mohanan, 2013). It is unclear how households respond to these shocks and whether they are capable of mobilizing resources to meet the increased medical costs and potential losses of income.

The main objective of this chapter is to estimate the household costs and the resulting coping strategies to face these costs. Our initial hypothesis is that given the lack of social security mechanisms and limited financial protection from NCMS coverage, health shocks are likely to lead to a double burden of costs arising from increased medical expenditure and loss of income. It is unclear whether a health shock will lead to sustained reduction in labour force supply (for the sick member); and if so, whether other household members might be able to increase their own working

time in order to compensate. As a result, the impact of health shocks on income is not clear. In addition, this present work also seeks to understand how households mitigate the consequences of a health shock by identifying potential 'coping strategies'. In this chapter, these coping strategies refer to the sale of assets, borrowing, remittances and reduction in savings. The research on coping strategies in the case of a health shock is not very well developed, especially in the context of China. Previous work from Liu (2016) has shown that investments in child's human capital formation and child labour are used as common strategy. Liu's work does not cover the outcomes under consideration in this chapter.

This chapter relies on a large-scale survey conducted in Ningxia to provide an estimation of the 'total immediate costs' associated with health shocks (in terms of medical expenditure and income losses), as well as a review of potential coping mechanisms. Differentials in response are also explored, based on the number of shocks experienced in the household, as well as the gender and position of the individual who experienced the shock (*e.g.* household head, working age member, elderly).

In this data, health shocks are fairly prevalent in all socio-economic and age groups. Significant increases in health expenditure associated with the shocks and impacts on household income are also detected (although only significant in the case of a health shock experienced by a household head or a male working age individual). It appears that households also meet the costs following a shock by borrowing from friends and relatives

or from a financial institution. Health shocks are associated with a clear increase in the probability to contract a debt and increase in the amount of outstanding debts, significant for all 'types' of shocks. In addition, in some cases, a reduction in household assets (including productive assets) and same year savings, and an increase in remittances is observed.

This chapter is organized as follow. Section 4.1 reviews the theoretical background that supports this thesis and Section 4.2 provides a summary of the methods and measurement. Section 4.3 presents the results for the analysis of medical expenditure, income and the four types of coping strategies for which data is available. Finally, Section 4.4 summarises the main conclusions.

4.1 Theoretical background

4.1.1. Loss of income

While the literature has put a lot of emphasis on the determinants and impacts of health expenditure on households' socio-economic outcomes (in particular in the context of universal health coverage), the evidence on 'indirect costs' of care is thinner (McIntyre D. , Thiede, Dahlgren, & Whitehead, 2006). In general, this term mostly refers to loss of income from loss of work time. This link between health and time spent on productive activities has been well established in previous literature. The founding works of Becker and Grossman on human capital (Becker, 1964; Grossman, 1972) served as a basis for this research question. In

particular, Grossman's works on the demand for health capital (1972) positions health as one of the main determinants of earnings and time allocation (on work and leisure) (see Chapter 6 for full description of these hypotheses).

As a result, at the individual level, health shocks should lead to a decrease in income resulting from a decrease in work time. Nonetheless, earnings are observed at the household level in this study. Other household members might be able to compensate for the loss of work time of the sick individual by increasing their own work time, thereby mitigating the potential negative impact of shocks on income.

4.1.2. Coping strategies

The extent to which different households protect their living standards by employing different coping strategies has also been investigated in the literature, although not particularly extensively in the case of health shocks. Alderman & Paxson (1992) broadly characterize coping strategies as ex-ante (risk management strategies to reduce the occurrence of shocks) or ex-post (strategies to mitigate adverse consequences following the occurrence of a risk).

In this chapter, only ex-post coping strategies are considered. There is also no exhaustive and widely accepted list of ex-post coping strategies. Coping strategies usually range from self-insurance through savings or acquisition of assets to be depleted in bad years (Dercon, 2002), use of formal or traditional credit institutions (Udry, 1994; Islam & Maitra,

2012); unearned income transfers within a family or in a community (group risk sharing) (Dercon, 2002; Sparrow, Hadiwidjaja, Yumna, Warda, & Suryahadi, 2014).

The evidence on the use of these coping strategies is mostly descriptive, and there is little theoretical basis to describe how households decide to employ a coping strategy over another. In general, scholars have shown that a set of factors will influence household's choice of a coping strategy. These factors include[4]: the nature of the shock and the magnitude of its impact on household welfare, as well as risk perception, socio-economic characteristics of the household (Berman, Quinn, & Paavola, 2013). In a framework analysing the choice of coping strategy, Heltberg, Hossain, Reva, & Turk (2012) classify coping strategies as behaviour-based, asset-based, or assistance-based. In their work, the type of capital (human, social, or financial) that are available to households (pre-shock) will affect the choices of uptake of coping strategies.

4.2. Model and measurement

4.2.1. Estimation strategy

Similar to other papers looking into health shocks, we estimate health shocks on a range of outcomes with the following specification:

$$Y_{ht} = \propto HS_{ht} + \beta X_{ht} + \gamma_{ht} + \varepsilon_{ht}$$

[4] This discussion is not limited to health shocks, but was applied to all types of shocks (idiosyncratic or common).

where Y_{ht} is the outcome of interest (*e.g.* household wage income, savings, loans) for household h and time t, X_{ht} a vector of time-variant household controls, HS_{ht} is a dummy variable which takes on the value of 1 if a household member in household h experiences a health shock, γ_{ht} the individual fixed effects, and ε_{ht} the error term. Following the strategy adopted in other similar studies, the impact of health shocks on those outcomes is investigated in the year of occurrence of the shock. Health shocks are included as exogenous regressors in this model specification. Standard errors are clustered at the village level to control for within-village error correlation. Because the model includes fixed effects, only time varying household socio-economic and demographic controls are included.

This empirical strategy eliminates the potential influence of time-invariant unobservable characteristics, as well as controls extensively for observed characteristics. The remaining source of potential bias could be linked to time-variant unobserved and uncontrolled for characteristics.

4.3.2. Dependent variables

Several measures are employed in this chapter to evaluate the immediate costs associated with health shocks and the corresponding coping strategies employed. The means, median, minimum, maximum, standard deviation and skewness of each of the dependent variables used in this chapter are presented in Table 4-1.

Medical expenditure and income data is used to evaluate the total costs of illness. Information on medical expenditure is collected in the consumption module in one single question. Household heads are also asked to declare the total household income (including financial gifts from family and remittances) in one question. This variable does not dissociate earned and unearned income, and is self-reported. In addition to this question on income, household heads are asked to report remittances from household members working away from home.

'Savings' corresponds to a self-reported variable that aims to measure the total savings made on a recall period of 12 months; meaning that there is no information on the total household savings for a given household. As a result, the analysis does not estimate whether depletion of savings accumulated in previous years is used as a coping strategy; which would have been a much more informative analysis, especially in the context of China where saving rates have been shown to be quite high.

Finally, households were also asked several questions with regards to borrowing. First of all, households were asked whether they borrowed any money in the past six months. They were also asked whether they had an outstanding loan, and the amount of the loan. It is not known whether the loans were taken from a financial institution or from friends, relatives or other village residents.

A section of the question intends to collect information on house characteristics (including sanitation and water access) and inventory of the household's durable goods and livestock in order to estimate the

wealth of the household [5]. For each category of good or house characteristic, a dummy variable was generated (equalling to one if the household possessed the characteristic or good in question; zero otherwise). One exception is the information on livestock, for which the number of livestock owned by the household was included (not a dummy variable). Weights are assigned to each item using principal component analysis.

The asset index A_h is calculated using the following method:

$$A_h = \sum k \left[f_k \frac{(a_{hk} - \bar{a}_k)}{S_k} \right]$$

where a_{hk} is the value of asset variable k for the household h, $\bar{a}_k$ its mean value across the sample, S_k its standard deviation and f_k the weights of the first principal component (O' Donnel, Van Doorslaer, Wagstaff, & Lindelow, 2008; Hafez, 2014). In addition to this asset index, wealth quintiles are generated by ranking households in each year and separating the distribution in five groups of equal size.

For each variable, the top 1% outliers were excluded (which might explain small differences in sample size between the variables). The presence of outliers can lead to deleterious distortion of estimates in a parametric setting (Osborne & Overbay, 2004). In these types of self-reported data, and given the relatively 'homogenous' population of interest (rural population in one province in China), outlier figures might be the result of mistakes in data entry, recall mistakes or misreporting. In the entire sample, a total of around a thousand observations were deleted for each

[5] This methodology was defined and presented in Hafez (2014).

variable of interest. For instance, on income, 830 observations were deleted, ranging from 102,000 to 500,000). Those households were not significantly from the mean household with regards to household size, marital status, education, number of migrants or gender of the household head. Nonetheless, they appeared to have a higher asset index score than the sample mean (0.64 versus 0.012) and higher consumption (86118 Yuan versus 51887 Yuan).

It is worth noting that the data used in the following analyses are mostly self-reported, which might be a source of concern as for their validity; including on income. Nonetheless, most household surveys rely on similar measures of wage income, as administrative source of data on income is not available in the vast majority of developing countries. Without the existence of a benchmark (e.g. tax records for a portion of the sample), verification and validity checks are impossible. Moreover, questions intended to document income in household surveys are typically less detailed than for consumption (which is also the case in this survey). Deaton (1997) shows that in India, where both types of variables are available, there are consistent discrepancies between income measured in survey and national income figures (although these latter might be also subject to important measurement errors).

While the problem is widely acknowledged and almost common to all surveys collecting information on income, there has been no solution in the literature. Some researchers avoid using information on income altogether and use consumption figures instead, judging them more reliable given the more detailed data collection process. In this study,

both variables of consumption and income are needed and addressed different aspects of the question. In Ningxia, household income is also earned through many different sources such as farming, grazing, fishing or other activities in the villages (craftsmanship, construction etc.). Therefore, there is no conceivable alternative measure that would collect more accurate information on income due to the diversity and largely informality of these income sources. Deaton (1992) underlines the difficulty to measure income in the developing world, especially in farming economies where income can usually be best understood as 'cash on hand'. In the absence of a possible correction, it was decided to use the variable whilst recognizing its limits and its almost definitive measurement error. One reassuring statistic is that household income and consumption show a strong and significant correlation of 0.41 ($p<0.01$). In addition, the correlation between income and asset index is also strong and significant (0.48); and stronger and more significant than the correlation between assets and household consumption (0.39). This might indicate that the measure of income is related, although imperfectly, to household's living standards.

4.2.3. Dependent variables

Health shock variable

Health shocks are defined as a hospital admission episode for causes that are acute, likely to be sudden (and therefore unplanned in timing) and to lead to long-term impairments. These include heart attacks, strokes, onset of a new cancer (except for skin cancer), diseases of the digestive

system, and accidents (see Chapter 3 for a full discussion of the health shock variable construction). Since this study only looks at household level outcome, the health shock variable equals to one if at least one household member has a health shock. A separate dummy variable is also created if more than one household member has experienced a shock, to investigate the impact of multiple shocks. The maximum observed number of shocks in a given household in our sample is 3 (in 27 households across 26803 household observations), and was declared in households with on average 6 members.

In addition, household differences in response and behaviour are also explored based on the status and gender of the individual who has experienced a shock. For a given household, the following health shock variables are defined:

- Health shock experienced by the household head
- Health shock experienced by any working age adult but the household head (also split between genders)
- Health shock experienced by an elderly (aged 55 or above)
- Severe health shock using a seven-day hospital admission threshold.

Previous literature has focused exclusively on the impact of health shocks experienced by household heads on household income (Lindelow & Wagstaff, 2005). This analysis also seeks to test whether there are differences in impact between household heads and other working age household members. We expect to see little impact of a shock experienced by an elderly on income. In this context only, an elderly is

defined as an individual aged 55 and above, in contrast to a working age individual. This is not the most commonly used threshold, but labour force participation after 55 years of age is rapidly decreasing in China (see Chapter 6).

Other covariates

These include household head's age, educational attainment, main occupation (decomposed in three dummies: farmer, labourer, unemployed); as well as the number of household members, number of migrant workers, number of individuals aged 55 and above (or 'seniors'), number of children, and an index that characterizes the quality of the house. The relationship between the household's age, educational attainment, number of children, labourer status, number of migrant workers and quality of the house with income and savings should be positive. Conversely, unemployment of household head, number of children and seniors (as they are less likely to contribute to productive activities) should have a negative relationship with these outcomes.

Table 4-1. Summary statistics on the dependent variables used in the chapter

Variable	Mean	Minimum	Maximum	Standard deviation	Skewness
Dependent variables					
Income	17,737	500	100,000	17,039	2.05
Medical expenditure	2,397	1	30,000	4,132	3.53
Asset Index	-0.65	-3.83	1.76	1.04	0.11
Loan (dummy)	0.63	0	1	0.48	-0.56
Loan (amount in Yuan)	15,962	0	200,000	27,290	3.31
Remittances	5,363	0	50,000	8,790	2.24
Savings	1,383	0	50,000	8,904	6.51

NB: the summary statistics are calculated for an average sample size of 25,688 observations in four years.

4.3. Results

4.3.1. Descriptive statistics

5.2% of the adult population experiences a health shock in a given year and this rate is quite stable across all four waves (although the rate is slightly higher -6.5%- in 2015). Only looking at individuals followed-up for the entire four waves (40.52% of the total surveyed sample), 13.2% of adults have experienced at least one shock in any given wave of the survey. Only 2% of adults have experienced more than one shock over the four waves. These results supports the definition of the health shock variable, as 82% of individuals who experienced a health shock over the four waves have only experienced no more than one health shock.

Health shocks also appear to be quite prevalent in all socio-economic groups. Table 4-2 provides a summary of the rate of shocks across age groups, wealth quintiles and education. Unsurprisingly (considering the definition applied), health shocks are more prevalent in after 35 years of age (and very prevalent after 65 years of age). A higher rate of shock is recorded in the lower quintiles and for adults with lower educational backgrounds.

Table 4-2. Rate of shock across wealth quintiles, age groups and education attainment

Variable	Rate of shock
Age groups	
16-24	1.48%
25-34	2.39%
35-44	3.97%
45-54	6.14%
55-64	10.01%
65+	14.80%
Asset quintile	
Quintile 1	6.57%
Quintile 2	5.39%
Quintile 3	5.17%
Quintile 4	4.67%
Quintile 5	4.66%
Education groups	
No school	8.96%
Elementary school	5.43%
Middle school	2.93%
High school or more	1.84%

Table 4-3 provides comparative statistics on individuals who have experienced a shock versus those who have not experienced a shock, and shows significant differences between the two groups. Individuals with a recorded health shock are usually older, with lower education, more likely to be farmers and less likely to migrate. Moreover, they are also more likely to live in a household with a slightly lower reported income and higher health expenditure in the preceding year. Some of these differences might be attributed to a cohort effect, as individuals who experienced a health shock are generally older. For instance, the share of young individuals who had received no education at all is usually lower

than for older individuals, as the law and programme to implement compulsory education was only implemented in 1986. Older individuals are also less likely to migrate.

Table 4-3. Differences in household and individual pre-shock characteristics between those exposed to shock and those not exposed

Pre-shock characteristics	Shock=0	Shock=1	Significance	P value (1)
Age	38.70	51.76	***	0.00
No education	25.91	46.03	***	0.00
% married	95.45	92.34	***	0.00
% farmer	61.57	72.35	***	0.00
% migrant	25.56	18.44	***	0.00
% household heads	30.67	34.92	***	0.00
Household size	4.48	4.52		0.20
asset index	-0.78	-0.85	*	0.08
Health expenditure	2,019	2,949	***	0.00
HH consumption	25,193	25,578		0.32
Wage income	15,054	14,203	***	0.00

(1) Obtained from Chi^2-test for categorical variables and t-test for continuous variables

*p <0.1, ** p<0.05, *** p<0.01

Many of these observed differences become non-significant when only considering adults aged between 45 and 60 in the sample (see Annex 4-1). The only remaining statistically significant differences in characteristics between individuals who experience a shock and those who did not are differences in education, household head status, wage income, household size and health expenditure. It is a caveat to keep in mind, as the expectation is that these shocks are sudden and unexpected and accordingly, there should be no differences in such measures. It is worth noting that all the regressions performed in this chapter control for these elements.

4.3.2. Regression estimates

Direct and indirect costs of health shocks

Results of fixed effects regression of household income are presented in Table 4-4. The impact of health shocks on income seems to differ depending on the identity of the household member experiencing a shock. This is hardly surprising, given the fact that household members do not contribute equally to the total labour force supply of the household. For instance, women tend to spend more time on household chores, care of members (including children) and non-wage labour (Liu, 2010). Health shocks experienced by household heads have a significant and negative impact on household wage income ($p<0.05$). It is worth noting that health shocks experienced by working age men (which includes household heads) is also significant, although only at the 10% level. The estimated reduction in income following a health shock experienced by a household head represents on average 6% of the total household income. These results are consistent but smaller than those of Lindelow & Wagstaff (2005), who estimated that a health shock experienced by a household head leads to a reduction of income ranging from 6.2% to 12.4% depending on the severity of the shock[6].

[6] Lindelow and Wagstaff (2005) use self-reported health to construct their health shock variable and construct two variables of health shock: large shock (drop of 2 or more points in self-rated health) and small shock (drop of 1 point in self-rated health). In their data, self-rated health is a five point scale.

The effects of health shocks experienced by other household members seem to be less clear. Interestingly, while insignificant, the estimated impact of a health shock experienced by an elderly seems to be positive (by contrast with those experienced by a working age adult). This positive sign could be an indication that other working age household members increase their time spent on productive activities in order to generate more income to meet the increased medical costs.

Health shocks are associated with a significant increase in medical expenditure for all types of shocks ($p<0.01$). Unsurprisingly, the largest estimated increases in medical expenditure are observed in household that have experienced several health shocks in the same year. In households with multiple shocks, the increase in medical expenditure is on average 2,624 Yuan. As a benchmark, the average medical expenditure across all four waves is 2,521 Yuan, which means that multiple health shocks on average more than double the medical expenditure in a given year.

Table 4-4. Results of regression of wage income and health expenditure (in Yuan)

Type of health shock	Wage Income	Health expenditure
Health shock (dummy)	271.5	1,827***
	(328.4)	(108.9)
Household head shock	-999.8**	2,044***
	(408.5)	(169.8)
Shock man	-948.8*	2,245***
(working age)	(515.6)	(211.1)
Shock woman	-193.0	1,607***
(working age)	(410.2)	(143.0)
Shock elderly	188.7	1,767***
	(396.1)	(142.0)
Severe shock	441.4	2,474***
	(383.0)	(150.6)
Multi-shock	-404.4	2,624***
	(933.1)	(414.3)
Mean of the dep. variable	17,737	2,397
Standard deviation of the dep. variable	17,039	4,132
Observations	25,240	26,375

Each column and line corresponds to a different regression. Since the categories are not always 'mutually exclusive' (e.g. shock experienced by a household head cannot be inserted along with shock experienced by a working age man since these two categories overlap). All regressions are performed with the following controls: education, age, marital status, gender and main occupation of household head, as well as house quality index, household size, number of migrants, children and seniors. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
Full regression tables in Appendix B (B1-B4)
*** p<0.01, ** p<0.05, * p<0.1

Other significant covariates include household head education, gender, main occupation (including unemployment), house index, number of migrants and children in the household and household size for income. For health expenditure, age and education of the household head, house

index, and number of seniors in the household are also significant covariates. Full regression tables are presented in Appendix B (B1-B4).

Coping strategies

The data available in the household survey can inform on some of the channels employed to meet the increased medical expenditure and losses of income. Table 4-5 shows the results for the impact of health shocks on household's twelve-month savings, remittances, assets and debts. The results are shown for any health shock experienced in the household, experienced by household heads, and for multiple household shocks.

These results show that to some extent, remittances are used to finance this type of emergency but the coefficient is not precisely estimated and only significant at the 10% level. The estimated coefficients correspond to approximately a 10% and 17% increase in remittances for a household member shock and a shock experienced by the household head. Most strikingly, health shocks seem to be associated with an increase in the probability of having a debt (not shown here) and amount of outstanding debts. Interestingly, the magnitude of the effect of health shock on debts is of a similar magnitude to the increase in health expenditure (presented in Table 4-3). This suggests that borrowing money is one of the main ways of coping with the costs associated with health shocks in China. 65.86% of households in our sample declared to have an outstanding loan, showing that access to such funds is quite widespread. There is little information on the source of the loan, *i.e.* whether the loans are taken with a formal financial institution or from a relative or friend.

This is also consistent with qualitative information collected from the questionnaire. In 2012 and 2015, 31.0% of households have declared taking up a loan to pay for medical care in the six months prior to the survey. In addition, household heads were asked each year to declare the primary reason for taking up a loan. 'Paying for medical care' was the most cited reason (32.41%), followed by 'building or repairing of the house' (19.4%) and by 'paying for wedding or funeral expenses' (12.76%).

Table 4-5. Regression results by coping mechanism (debts and savings are in Yuan, Assets refers to the Asset Index)

Type of health shock	Debts	Remittances	Savings	Assets
Health shock (dummy)	2,100***	567.6*	-144.8**	-0.00
	(476.2)	(328.0)	(66.5)	(0.00)
Household head shock	1,662**	939.6*	-190.4**	-0.03**
	(644.5)	(498.0)	(72.9)	(0.01)
Multi-shock	4,476***	1,025	-142.1	0.04
	(1,788)	(905.7)	(179.1)	(0.03)
Mean of the dep. variable	15,862	5,363	1,383	-.65
Standard deviation of the dep. variable	27,290	8,790	8,904	1.04
Observations	26,539	12,995	25,544	26,418

Each column and line corresponds to a different regression. Regressions on remittances are only performed on households with at least one migrant worker. All regressions are performed with the following controls: education, age, marital status, gender and main occupation of household head, as well as house quality index, household size, number of migrants, children and seniors. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
Full regression tables in Appendix B (B1-B4).
*** p<0.01, ** p<0.05, * p<0.1

In addition to the estimated impact on loans, households tend to save less as a result of a shock in the previous 12 months (although not

significant for multiple shocks). Again, this variable does not show whether households use their accumulated savings to pay for medical expenses as it only refers to savings accumulated in the preceding year. It is worth noting that on average only 14% of individuals in this sample declare to save money every year. This rate is not too different from some of the saving rates put forward in the literature for low-income households. According to Yang, Zhang, & Zhou (2011), low income households are not very likely to save money (only 7% did in 2007) compared to high-income households (34% in 2007). Finally, a 4% decrease of the household assets is detected following a health shock experienced by a household head ($p<0.05$), not significant for the dummy variable of shock and for multiple shocks.

Differences in coping strategies based on asset quintile, education of the household head and food consumption (poor consumption defined as a reported food consumption below the community median) are further explored separately (for health shock dummy at the level of the household). There is a consistent pattern observed using this decomposition. The first conclusion is that regardless of the group considered, the impact of health shock on borrowing is always significant, for all groups. Reduction in savings and sale of assets is only significant for households with lower education and for households considered poor with regards to food consumption. Results are not reported but available on request.

4.4. Conclusion and discussion

In recent years, the health economics research has almost exclusively focused on medical expenditure to look at the financial burden associated with health problems. This approach has been incomplete, as medical costs do not cover the full economic burden experienced by households. This analysis contributes to the existing literature by estimating the immediate costs of shocks and reviewing the types of coping strategies employed by households.

This in-depth analysis of the household response to health shocks has highlighted several important elements. First of all, households in the province of Ningxia experience a double economic burden following a health shock in the form of a decrease in income (significant for household heads only) and increase in medical expenditure in the year of the shock. The increase in medical expenditure is of a larger magnitude than the loss of income. This highlights potential compensation strategies from other household members, including increasing labour force supply of spouses and other working age members. The analysis of labour force distribution within the household will be provided in Chapter 6.

This data only allowed the investigation of four different types of coping strategies: remittances, loans, sale of assets and savings. Firstly, consistent and significant increases in loans are shown for all types of shocks, regardless of the position and identity of the household member who experienced the shock. The magnitude of the estimates matched

quite closely the measured increase in medical expenditure following a shock. This suggests that borrowing from formal sources (e.g. a financial institution) or friends and relatives is an important channel for households. Moreover, health shocks have a negative impact on same-year savings. Nonetheless, this variable only refers to past year saving (with a recall period of 12 months). Interestingly, shocks experienced by the household head are associated with a significant reduction in assets (ranging around 4% of the household assets). Finally, for households with at least one migrant worker, we observe that health shocks lead to some increase in remittances, although this result is only significant at the 10% level.

This analysis suffers from several shortcomings. First of all, most of this data is self-reported and there is no source for validating the declared information. This is a particular concern for the income variable, which most likely suffers from reporting problems. As highlighted by previous researchers, there might be a misunderstanding from households on the definition of what constitutes income (Deaton, 1997). This is a problem encountered by almost all studies relying on survey data, especially in the developing world; and there is no widely accepted solution to it (authors sometimes do not mention this serious limitation). The interpretation of the results was made in knowledge of this limitation. Moreover, while the definition of the health shock variable was made with the support of existing literature on the topic, comparison in pre-shock outcomes showed significant higher medical expenditure for households that experienced a health shock. This poses a potential threat to the exogeneity assumption of the health shock variable. Concerns on this

point are addressed by applying individual fixed effects and controlling for a broad set of covariates that might influence these baseline differences in means.

An extremely large share of households (65.8%) reported outstanding loans. This is concerning, given the very high average loan (15,934 Yuan per household), and the potential high level of indebtedness of households for medical reasons. In 2015, 98.5% of the study population was covered by the NCMS; yet the analyses from this chapter show that health shocks consistently lead to important increases in medical expenditure, and in some cases, with significant negative effects on household income, savings and debts. This calls on more research on how to provide an effective safety net using both social protection mechanisms and existing health insurance.

Annex 4-1. Differences between individuals who have received a shock and those without a shock, restricting the sample to individuals aged between 45 and 60

The following table presents the differences in baseline characteristics (pre-shock) between individuals who have experienced a shock versus those who have not (not standardised). Compared to the results presented in Table 4-3 (on a non-restricted sample), we observe fewer differences and differences of a smaller magnitude. As argued in throughout this chapter, these differences could be explained by cohort effects as individuals who have experienced a shock are on average older than those who have not experienced a shock.

Pre-shock characteristics	Shock=0	Shock=1	Significance	P-value (1)
Age	50.2	50.9	***	0.00
No education	36.7	46.4	***	0.00
% married	95.1	94.6		0.69
% farmer	84.0	86.9		0.21
% migrant	11.9	10.1		0.66
% household heads	51.8	41.0	***	0.00
Household size	4.51	4.37	***	0.00
asset index	-0.69	-0.72		0.41
Health expenditure	2,351	3,025	***	0.00
HH consumption	28,361	28,045		0.70
Wage income	17,035	15,302	**	0.01

(1) Chi^2-test performed on categorical variables, t-test on continuous variables

*p <0.1, ** p<0.05, *** p<0.01

Restricting the sample to individuals aged 45-60 years of age, individuals who experienced a health shock are less likely to be household heads, more likely to live in smaller households, have lower education, report higher health expenditure in the preceding year, and lower income. Differences in wage income corresponds to about 11%. It is worth noting that this measure of wage income variable is not adjusted to the number of household members. If adjusting the wage income per capita (only

considering adults), the difference is smaller (ranging around 7%) and only significant at the 10% level (not shown). On the other hand, there remain important differences in health expenditure of even after adjusting for differences in household size.

Chapter 5. The impact of health shocks on consumption: is consumption smoothing possible?

5.1. Introduction

As shown in Chapter 2, the first references to the term 'health shock' is found in studies risk and insurance, in particular on the testing of the full insurance theory (Gertler & Gruber, 2002; Morduch, 1995; Islam & Maitra, 2012). These latter intend to estimate the impact of an idiosyncratic shock to income (by contrast to common risks such as weather events, political crises or conflicts) on household welfare. Their objective is to test the completeness of insurance mechanisms in a community in the absence of perfect credit markets; thereby assessing the value and potential gains stemming from social insurance mechanisms (Chetty & Looney, 2006). According to this literature, the fact that households are able to avoid large shortfalls in consumption following a shock is in itself evidence that self-insurance or informal insurance mechanisms in communities are effective and that the gains from implementation of social insurance will be minimal (Morduch, 1995). In this framework, consumption is used a proxy for household welfare.

In Chapter 4, health shocks were shown to lead to an important increase in medical expenditure and losses of income (although not significant for all household members). Following the literature on the full insurance hypothesis, the extent to which this 'double burden of cost' impacts household welfare can be first analysed through the lens of consumption patterns. In this present chapter, we estimate the impact of health shocks on different categories of consumption (total, food, non-food non-medical) and at different points in time (in subsequent wave, and up to 6 years after the occurrence of the shock). The growing literature on shocks shows that households might be able to avoid shortfalls in consumption following a shock but that the uptake of inefficient coping strategies will be responsible for a deterioration of household welfare in the long-run (Chetty & Looney, 2006). The aim of this chapter is also to explore this hypothesis by estimating the impact of health shocks on consumption several years after the occurrence of the shock.

Besides the work of Liu (2016), there are no studies looking at the impact of health shocks on consumption patterns in China. Liu (2016) simply states that households are able to smooth consumption following a health shock, but the results are not presented. Nonetheless, Liu's study uses a very different measure of health shock to this study (*i.e.* 'number of days unable to carry out productive activities in the last 30 days'), which might lead to some biases in estimates. This measure of health shocks is very generic and it is unclear what the cause of this reduction in productive activities is. From previous studies in other research settings, the impact of such idiosyncratic shocks on consumption is not clear. Some studies have estimated a significant (and sometimes very large)

negative impact on consumption (Gertler & Gruber, 2002; Asfaw & Von Braun, 2004; Wagstaff, 2007; Gustafsson-Wright, Janssens, & van der Gaag, 2010). Other studies have shown that households were able to absorb the costs associated with the shock (although without much discussion on the mechanisms whereby consumption smoothing is achieved) and therefore found no significant impact of shocks on consumption (Powell-Jackson & Hoque, 2012; Mohanan, 2013). These studies were set mostly in developing countries. Chapter 2 provides a full review of this literature.

The material in this chapter is organized as follow. Firstly, Section 5.2 presents the data and methods used in this chapter, and Section 5.3 the results of the analysis. Finally, Section 5.4 discusses the results and conclusions from this work.

5.2. Data and methods

5.2.1. Methods

The impact of health shocks on consumption is investigated in two different steps in order to identify the short-term and long-term impacts.

Firstly, we estimate the following equation:

$$Y_{h,t} = \propto HS_{h,t} + \beta X_{h,t} + \gamma_{h,t} + \varepsilon_{h,t}$$

where $Y_{h,t}$ is the same-year household consumption (total, food or non-medical non-food or NMNF) of household h and time t, $X_{h,t}$ a vector of

time-variant household controls, $HS_{h,t}$ is a dummy variable which takes on the value of one if a household member in household h experiences a health shock and zero otherwise, $\gamma_{h,t}$ the household fixed effects, and $\varepsilon_{h,t}$ the error term. Health shocks are included as exogenous regressors in this model specification.

The long-term impacts of health shocks on consumption patterns are estimated using a first difference equation (for household heads only) as follows:

$$Y_{h,t} - Y_{h,0} = \propto HS_{h,0} + \beta(X_{h,t} - X_{h,o}) + \varepsilon_{h,t}$$

which follows the notation presented above, and where t here is the number of years since the occurrence of the shock, or in other words t = 0, 2, 3, and 6 for respectively 2009, 2011, 2012 and 2015. In this specification, $HS_{h,0}$ is the health shock experienced in 2009 (t=0). At t=0, we simply do a cross-sectional regression of the impact of health shocks on same year consumption.

The above first difference regression intends to explore differences in consumption patterns of households that declared to have experienced a shock in the 2009 wave (treatment) compared to those of households that did not experience a shock in 2009 (control) at different points of time after the shock. In other words, this strategy enables us to measure the impact of health shocks after two, three and six years using a specification similar to a difference in difference equation where treatment is experienced in 2009.

It is worth noting that post-2009 health shocks are not controlled for and that some households in both the control and the treatment group experience a health shock in subsequent periods. An alternative strategy could be to only include households that have not experienced a health shock in any of the waves (about 4,200 households, or 60% of the sample of households) in the 'control group'. While both options are explored, this option was discarded. The exclusion of households that experienced subsequent health shock did not change the results, although the magnitude of the estimated impacts is on average 15% larger than those reported in the following section (for all types of consumption). We observe no change in the sign of the estimated coefficients or significance of the results.

A similar strategy can be found in García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer (2013), which investigates the impact of health shocks on individual long-term employment status and income in the Netherlands. Similarly, the authors do not control for subsequent health shock occurrences in both the control and treatment group.

All regressions include the following regressors: house quality index, number of migrants in the household, household size, number of children and elderly, as well as household head educational attainment, marital status, occupational attainment, gender and age. House quality index is included as a general indicator of household's wealth, but was chosen over household assets in order to avoid endogeneity problems. House quality index, household size, household head educational attainment, marital status, and age are expected to be positively

correlated with all three consumption categories. Controlling for household size, we hypothesize that the number of children and elderly will have a negative association to consumption.

5.2.2. Data and health shock variable

The data used in this chapter is the Ningxia project household survey. All four waves (2009, 2011, 2012 and 2015) were used for this analysis. The health shock variable used in this chapter is also described in Chapter 3 and descriptive statistics pertaining to the characteristics of individuals and households that experienced a health shock are reported in Chapter 4 (Section 4.3.2). Similarly to Chapter 4, heterogeneity of impact based on health shock characteristics is also investigated using the following categories of health shocks:

- Health shock (dummy, for any household member)
- Health shock experienced by the household head
- Severe health shock (hospitalization for more than 7 days, for any household member)
- Multiple health shocks within the same household

The long-term impacts are presented only for household head health shocks using all four waves. This choice was taken for presentational purposes and household heads were chosen as those shocks have been shown to be lead to important disruption in the previous chapter (compared to other household members).

Out of 6701 households observed in 2009, respectively 81.1%, 76.9% and 63.7% were successfully followed-up in 2011, 2012 and 2014 (see Chapter 3 for discussion on attrition).

5.2.3. Consumption figures

The accompanying household survey to the Ningxia Project contains detailed questions on the consumption of individual items, which are then aggregated to calculate the total household consumption in a given year. The consumption-reporting module is split in two. For a first set of consumption categories (*e.g.* meat, vegetables, tobacco, transportation, commodities, rent, cultural activities), households are asked to report their monthly expenditure using the last month as a recall period. In a second part, households are then asked to report expenditure in Yuan for a set of goods and activities using the last 12 months as a recall period. This section collects information on large events (*e.g.* funeral, wedding), house construction expenditure, health expenditure, or education.

In addition to the food expenditure reported, households are asked to declare consumption of home produced food for ten categories of goods (*e.g.* eggs, fruits, grains, meat). Quantities of home produced food are multiplied by the self-reported market price and yearly food production is calculated by multiplying the collected data by twelve. The value of the total yearly home-produced food consumption is added to the reported expenditure on food collected as described above in order to construct the variable of yearly food consumption.

Total household consumption is obtained by adding together the categories presented above (including home-produced food) for a given household. In addition, non-food non-medical (NMNF) consumption is also obtained by subtracting food and medical consumption from total household consumption. This category of consumption includes the following types of consumption: daily commodities, utility bills, rent, cigarettes, alcohol, transportation, phone, education, clothing, gifts, entertainment (including gambling), taxes, cultural activities, purchase of non-durable goods, building and or repairing buildings.

Similarly to the variables used in Chapter 4, the top 1% outliers for each variable and zeros are treated as excluded from the sample (on average 2000 observations per category of consumption).

Table 5-1 provides a summary of the minimum, maximum, mean, median and skewness of each of category of consumption (without the top 1% outliers).

Table 5-1. Summary statistics of the consumption variables (in Yuan, not adjusted)

Variable		Mean	Minimum	Maximum	Standard deviation	Skewness
Total household consumption		30,471	600	283,400	26,718	2.86
Averages	2009	19,467	1,058	137,680	13,649	2.43
	2011	26,907	2,710	242,310	21,116	2.89
	2012	31,353	600	235,097	25,703	2.84
	2015	44,107	650	283,400	35,233	2.14
Food consumption		7,319	584	27,880	4,838	1.29
NFNM consumption		18,803	1,010	133,540	19,366	2.48

NB: the summary statistics are calculated for an average sample size of 26,800 household observation (6,500 households surveyed in each year).

5.3. Results[7]

5.3.1. Average impact over all four waves

Health shocks lead to an increase in total household consumption, regardless of the type of shock or identity of the individual who experienced the shock (*e.g.* whether the shock was experienced by the household head) (all results significant at the 1% level) (Table 5-2). This increase in total household consumption is mainly driven by an increase in medical expenditure (see Chapter 4), and is, unsurprisingly larger for households that experienced severe health shocks or multiple health shocks. It is worth noting that the increase in total household consumption is of a similar magnitude for the health shock experienced by a household head and the simple health shock dummy at the level of the household.

However, if now only considering food consumption, health shocks experienced by a household head are associated with a significant decrease in consumption ($p<0.05$). This corresponds on average to a 4% decrease in the total food consumption (Table 5-2). By contrast, no significant impact was estimated for other types of shocks. Previous studies on health shocks did not explore these differences based on the identity of the member who experienced the health shock. One explanation for this observed difference between household heads and

[7] For descriptive information on health shocks and its association with socio-economic variables, please refer to Chapter 4 (Section 5.4.1).

other household members could be explained by the fact that the household head is not only the primary breadwinner but also usually the central planner within the household (Dercon & Krishnan, 2000). In this sense, health shocks experienced by household heads might lead to shortfalls in consumption, as household heads are usually the primary decision-maker with regards to coping strategies.

Table 5-2. Regression results by types of consumption (in Yuan)

Type of health shock	Total consumption	Food consumption	NMNF consumption
Health shock (dummy)	2,551***	69.7	141.8
	(516.6)	(105.4)	(363.6)
Household head shock	2,542***	-246.1**	-348.7
	(694.6)	(119.0)	(495.7)
Severe shock	4,492***	96.8	937.2**
	(579.3)	(113.5)	(430.0)
Multi-shock	4,825***	269.6	1,983
	(1,591)	(274.2)	(1,228)
Observations	25,065	26,065	26,067
Mean of the dep. variable	30471	7319	18803

Each column and line corresponds to a different regression (individual fixed effects). All regressions are performed with the following controls: education, age, marital status, gender and main occupation of household head, as well as house quality index, household size, number of migrants, children and seniors. Differences in sample size between regressions might arise from missing values or removal of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
Full regression tables in Appendix B (B5-B11)
*** p<0.01, ** p<0.05, * p<0.1

It is also worth noting that severe shocks are associated with a significant increase in non-medical non-food consumption. This is a puzzling result as this increase is not caused by changes in food and health expenditure. We explore this question by further estimating the impact of severe shocks on more detailed categories of consumption (*e.g.* rent,

commodities, construction, gambling) (results not reported in this manuscript but available on request). It appears that severe shocks are significantly associated with an increase in gifts expenditure (of 410 Yuan) ($p<0.05$). It remains unclear what this increase in gifts corresponds to, although informal payments to health care workers have been reported in other studies (Eggleston, et al., 2007).

5.3.2. Long-term impacts

The results of the first difference equation exploring the long-term impacts are presented in Table 5-3 and 5-4, for total consumption, health expenditure, food consumption, NFNM consumption.

In 2009, 7.2% of households have reported a health shock experienced by the household head. Each column corresponds to the regression estimates obtained using a first difference specification (one analysis per column), and 2009 as a baseline. It is worth noting that the first column corresponds to a cross-sectional equation on the impact of health shocks on same year consumption (and is therefore not entirely comparable to the subsequent columns). Nonetheless, the patterns of results do closely match those put forward in the fixed-effects regression, although the magnitude of effects are different.

As shown in Table 5-4, health shocks are associated with an increase in total consumption and health expenditure in the same year (in line with the results shown above). Nonetheless, it is worth noting that 2009 health shocks are associated with large scale reductions in total

household consumption after 2, 3 and 6 years. These reductions are particularly important after six years, as the estimated effect after six years corresponds to a 24.1% reduction in total household consumption. Health shocks are associated with an increase in health expenditure, only significant in the year of the shock and 2 years after the shock.

Table 5-3. Long-term effects by type of consumption (total consumption and health expenditure)

Years since health shock	0 years	2 years	3 years	6 years
Total consumption	2,795***	-4,479***	-4,773***	-7,672***
	(663.7)	(1,377)	(1,383)	(1,922)
Observations	6,525	5,405	5,133	4,245
Mean of the dep. variable	19,467	23,168	25,478	31,785
Health expenditure	2,388***	1,379***	605.2	600.6
	(228.8)	(305.9)	(378.1)	(466.7)
Observations	6,665	5,426	5,140	4,262
Mean of the dep. variable	1,832	1,942	2,129	2,553

Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). All regressions are performed with the following controls: education, age, marital status, gender and main occupation of household head, as well as house quality index, household size, number of migrants, children and seniors. Differences in sample size between regressions might arise from missing values or removal of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
Full regression tables in Appendix B (B5-B11)
*** p<0.01, ** p<0.05, * p<0.1

Decomposing the effects using two categories of consumption (food and NFNM consumption), again, we find that health shocks have a significant negative impact on both food and NFNM consumption after 2, 3 and 6 years. The impact of health shocks on both type of consumption in the same year is not significant (although the impact on food consumption is

significant when using fixed effect estimation and all four years of data). The effects are very large, especially after six years. Health shocks are respectively associated with a 20.0% and 34.1% decrease in food and NFNM consumption after six years. The impact of health shock was estimated using a full set of household characteristics (and household head characteristics) and this large estimated effect cannot be accounted for by other observed changes in the households such as changes in household size (*e.g.* death of the individual who experienced the shock), occupational status of the household head, etc.

Table 5-4. Long-term effects by type of consumption (food consumption and NFNM consumption)

Years since health shock	0 years	2 years	3 years	6 years
Food consumption	112.6	-1,023***	-1,173***	-1,431***
	(170.3)	(323.1)	(297.3)	(361.5)
Observations	6,545	5,420	5,139	4,245
Mean of the dep. variable	5,741	6,514	6,771	7,093
NFNM consumption	338.7	-4,122***	-4,325***	-6,889***
	(562.1)	(1,031)	(1,045)	(1,457)
Observations	6,570	5,414	5,134	4,252
Mean of the dep. variable	11,709	14,712	15,482	19,671

Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). All regressions are performed with the following controls: education, age, marital status, gender and main occupation of household head, as well as house quality index, household size, number of migrants, children and seniors. Differences in sample size between regressions might arise from missing values or removal of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
Full regression tables in Appendix B (B5-B11).
*** p<0.01, ** p<0.05, * p<0.1

5.4. Conclusions

In line with other studies on the full insurance theory and health shocks, the analyses presented in this chapter provide evidence of significant negative impacts of shocks on food consumption (Gertler & Gruber, 2002; Wagstaff, 2007; Powell-Jackson & Hoque, 2012). It is worth noting that this negative impact is not significant for all types of health shocks, only for health shocks experienced by the household head. Other studies in the past have focused on health shocks experienced by household heads only, as they are presumed to have larger effects on households than for other household members (Dercon & Krishna, 2002). Nonetheless, one of the striking lessons from this chapter is that health shocks might have delayed adverse consequences on household welfare. All of the reviewed studies on the full insurance theory have focused on the consequences of health shocks on same year consumption figures, and therefore focused on some of the more immediate consequences to households. Chetty & Looney (2006) show that households might be very risk adverse and therefore employ any strategy to avoid shortfalls in consumption. These coping efforts are especially more important in poorer households, where a reduction of consumption will have a greater welfare cost. For these reasons, consumption levels are increasingly considered an imperfect proxy for household welfare. The analyses carried out in this chapter show that adopting a larger time frame (up to six years) could help shed light on the impoverishing effects of poor coping choices.

While health shocks experienced by a household head have an impact on the health expenditure in the same year and two years later (although of a

lesser magnitude), we found that such impact is not persistent in time (not significant three and six years after the shock). Nonetheless, households report lower total consumption in subsequent years. This result is driven by very large reductions in food and NFNM consumption, by far surpassing the 'immediate' recorded shortfalls in same year consumption. As these effects are not necessarily driven by an increase in health expenditure, we hypothesize that other channels, such as a long-term reduction in income (*e.g.* following the development of a disability) or inefficient coping strategies such as high indebtedness of households or sale of agricultural assets, are responsible for these trailing long-term effects.

Chapter 6. The impact of health shocks on labour force supply: a household approach

Estimations presented in Chapter 4 showed that health shocks experienced by household heads lead to a decrease in total household income (on average 5%). However, this effect is not significant for health shocks experienced by other household members. These results suggest that, in some cases, intra-household transfers of labour might help prevent shortfalls in income when shocks occur. While health shocks are expected to have a negative impact on own labour force supply, it is unclear what the labour supply response of other household members is. Compensation from other members was identified as an important coping strategy in previous works (Bonfrer & Gustafsson-Wright, 2016; Sparrow, Van de Poel, Hadiwidjaja, Yumna, Warda, & Suryahadi, 2014). On the other hand, spouses and other household members can decrease their labour force participation in order to care for an individual who experienced a health shock.

Previous studies have investigated this relationship between health shocks and own labour supply, although a large portion of this literature is set in high-income countries. Numerous studies showed that health shocks have a negative impact on labour force supply, increasing the

probability of labour force exit (Riphahn, 1999; Coile C. C., 2004; Jiménez-Martín, Labeaga, & Vilaplana Prieto, 2006), of early retirement (Hagan, Jones, & Rice, 2008), and of claiming disability or unemployment benefits (García-Gómez, 2011). All the above studies also emphasize the heterogeneity in response based on contextual factors, such as the generosity of unemployment and disability benefits, differences in retirement policies, and insurance status.

This research has received less attention in LMICs. In the absence of well functioning social safety nets, loss of work time following a shock will often translate into loss of income for households. Gertler & Gruber (2002) estimate that in Indonesia, health shocks have a very large negative impact on labour force supply, as severe health shocks increase the probability of labour force exit by 75% and decrease hours of work by 84%. In China, two studies from Lindelow & Wagstaff (2005), and Liu (2010) find that health shocks have a negative impact on own hours of work (although only significant for men and older workers in Liu's work). Liu (2010) also shows that women increase their work hours when their spouse experiences a health shock but that this 'compensation' is not sufficient to avoid drops in total hours of work in the household. Lindelow & Wagstaff (2005) find no evidence of intra-household labour force transfers.

The objective of this chapter is to investigate the loss of work time following a health shock for the individual who experienced the shock and the rest of the household (including spouses). This chapter relies on the Chinese Family Panel Studies, and uses first difference regressions

and inverse probability of treatment weighting (based on propensity scores) to control for time-invariant characteristics and differences in baseline characteristics. By contrast to the work conducted by Liu (2011), different measures of health shock and labour force supply are used, and all household members are considered (not only spouses). Heterogeneity in response based on household and individual characteristics is also explored. For instance, are there differences in response conditional on individual's work status at baseline, or based on the relative position in the household (*e.g.* whether individual is a household head), age or type of residency (urban or rural)?

Analysis in this chapter show that health shocks have a significant negative impact on work hours, stronger for women than men and only significant in rural areas and for older adults. Health shocks experienced by a spouse leads to a significant increase in labour force supply of wives (not significant for men) and older workers only. Finally, no significant impact on the labour force supply of other household members (excluding spouses) were detected.

This chapter is organized as follow. Section 6.1 provides background and theoretical basis for this analysis. Section 6.2 discusses methods and describes the variables used in this study. The results of the analyses conducted on health shocks and labour force supply are presented in Section 6.3. Finally, conclusions and discussion of limitations are provided in Section 6.4.

6.1. Background and theoretical basis

The literature on the impact of health shock on work time and labour force participation has been very prolific in recent years, in both LMICs and high-income countries. Most of this literature relies on the seminal works of Grossman (1972) on the modelling of health as a human capital, which offers a good starting point to discuss the impact of health shocks on labour force supply. In Grossman's model (1972), a given individual's time is divided between productive and non-productive time, time spent on production of health inputs and time spent on sickness. The utility derived from the time spent on sickness is zero. Time investments dedicated to the production of health lead to lower time spent on illness, and therefore increase the time spent on productive and non-productive (*e.g.* leisure) time. Individual's productive and non-productive time is valued at the market wage rate. Other authors have derived other pathways through which poor health can impact time allocation from Grossman's model, including through its impact of productivity or other forms of human capital acquisition.

In this framework, health shocks can have a negative effect on hours of work through the following channels:

(a) Health shocks can lower overall time dedicated to productive and non-productive activities by increasing the time spent ill or time spent on production of health (*e.g.* accessing health services)

(b) Health shocks can also impact work hours by reducing productivity, thereby reducing market wage rate. In this case, two possible effects can be observed:

- A **substitution effect**: as stated above, individuals value their productive and non-productive time at market wage rate and allocate their time on each type of activity based on their preferences. A change in wage rate can change the time allocation trade-off and result in individuals spending less time on productive activities and increase non-productive time if they derive higher utility from it.
- An **income effect**: an individual's purchasing power increases with their income. As a result, if health shocks lead to a decline in individual's market wage rate, they might also result in an increase in work hours in order to avoid reductions in income and purchasing power.

The first channel (labelled (a)) will undoubtedly lead to a decrease in work hours. Nonetheless, if health shocks have lasting effects on individual's wage rate (channel (b)), the impact on work hours is not clear. The magnitude of substitution and income effects will depend on individual's preferences and on the severity of the shock. The income effect might also be 'exacerbated' by the lack of safety nets (*e.g.* disability benefits, unemployment insurance).

The Grossman model (1972) illustrates some of the time allocation trade-offs for an individual but gives little insights on the trade-offs within a household and the labour response from other household members. As highlighted by Gallipoli & Turner (2011), intra-household risk sharing and insurance is the primary mechanism to protect individuals from experiencing the adverse consequences of an idiosyncratic shock. Many

household models have been formulated in recent years, with very different assumptions (*e.g.* Pareto versus non-Pareto, bargaining versus cooperative) and modelling applications (Vermeulen, 2002). Cooperative models are the most frequently used models, and assumes that while household members have distinct preferences, these latter are modelled as a function of own consumption and consumption of other household members. Cooperative models also postulate that income is pooled in the household.

Without entering into too much depth in the details of these models, their premises show that the household offers the most immediate source of risk-sharing between individuals. Applied to health shocks, one natural prediction from this generic model is that health shocks will lead to transfers of labour between healthy and unhealthy household members, in order to mitigate the negative impacts of shocks on income and consumption. Such substitution of labour can come from one or several member in the household, and the additional 'contribution' of each individual reflects his/her position within the household, initial relative contribution (in the absence of a shock). The existence of other support mechanisms other than labour force substitution (*e.g.* insurance) and other types of external constraints (*e.g.* labour regulations) can also affect the response from other household members. In this sense, a decomposition of this analysis by individual characteristics and urban or rural context might provide information on this heterogeneity in response.

It is worth noting that other forms of support within the household can be provided, such as informal care. This latter might lead a decrease in work hours supplied by healthy household members following a shock.

6.2. Methods and data

6.2.1. Empirical strategy

As stated above, this chapter explores the relationship between health shocks and own labour force supply, as well as labour supply of other household members (in particular spouses).

First of all, the following relationship is estimated:

$$\Delta WH_{it} = \propto \Delta HS_{it} + \beta \Delta X_{it} + \varepsilon_i$$

where WH_{it} is the formal work hours of individual i and time t, HS_{it} is a dummy variable for whether the individual has experienced a health shock, X_{it} is a set of time variant covariates, and ε_i is the error term.

In the same manner, the household effects on other members who did not experience a health shock are estimated using the following specification:

$$\Delta WH_{it} = \propto \Delta HS_{ht} + \beta \Delta X_{it} + \varepsilon_i$$

which follows the same notation as above, but in this case, the term HS_{ht} refers to whether anyone in the household has experienced a health shock. In a first analysis, we restrict the sample of individuals to couples and investigate the spousal labour response to a health shock. We

postulate that spousal response (in terms of formal work hours) will be stronger than response from other household members. Several empirical works have shown that the labour force response to spousal health shocks can be large and significant (Coile, 2004; Liu, 2011). In a second step, we extend this analysis and investigate the response of any other individual in the household. The provision of informal care cannot be studied in this analysis, as only information on formal work hours is collected in the survey. Nonetheless, if we observe a sharp reduction in formal work hours from spouses or other household members, one potential explanation could be that those individuals are reducing their labour force supply in order to care for the sick household member.

The dependent variable (work hours) is kept in levels as opposed to a logarithmic transformation, as we are interested in the total effect of health shocks on work hours, and keeping it in levels will also help us assess the total effect at the household level (Liu J. X., 2010). In addition, changes in labour market participation (whether an individual has declared a positive number of work hours in the last twelve months) are also estimated using a linear probability model.

The above regressions are estimated using first difference estimation, combined with inverse probability of treatment weighting (IPTW) using propensity scores. For each individual, a propensity score is calculated as the probability of receiving treatment (*i.e.* a health shock, spousal health shock, or household shock) conditional on observed baseline characteristics. Rosenbaum and Rubin (1983) define propensity scores as

the probability of treatment assignment conditional on observed baseline characteristics, or

$$\pi_i = \Pr(t_i = 1 \,|X_i)$$

where π_i is the estimated probability of treatment assignment t_i, and X_i a set of observable characteristics. Separate propensity scores are calculated for each of the groups of interest, therefore estimating the following:

- The probability of experiencing a health shock based on a series of baseline observable characteristics
- The probability of being married to an individual who has experienced a health shock based on baseline observable characteristics
- The probability of living in a household where an individual has experienced a health shock based on baseline household characteristics.

For the first two types of propensity scores[8], individual characteristics that have been identified as predictors of both health shocks (spousal and own) and labour force supply are included, as advised in the literature (Curtis et al., 2007). These include the following characteristics at baseline:

- Individual characteristics: age, sex, marital status, education, whether the individual is Han, whether the individual is the household head

[8] The propensity score for experiencing a health shock in the household (excluding spousal health shock) was calculated using a reduced set out variables, to satisfy the balancing property. These only included household level characteristics as well as age, education.

- Household characteristics: household size, number of working age individuals, urban dummy, whether household consumption is below median in the county, access to running clean water in the household.

Propensity scores are estimated using a logistic model. Unsurprisingly (given the size of the total sample and the relatively few cases of health shock), there is evidence of common support in our sample. Common support ensures that there is a strong overlap in estimated propensity scores in the group of those who experienced a health shock and those who did not. This condition is necessary to properly apply propensity score based techniques, as it ensures that the sample contains individuals with similar probability of experiencing a health shock, but with different health shock status (or 'treatment status'). Please see Annex 6-1, which plots the respective distribution of propensity scores across both groups. Both propensity scores for health shock and spousal health shocks are comprised between 0 and 0.20. Propensity score for household shock is comprised between 0 and 0.36.

The balancing property for all three types of propensity scores is also satisfied using these predictors. Balancing of properties is also central to the use of propensity scores, as it ensures that the propensity score is properly calculated and that its main purpose (*i.e.* to balance properties of both groups) is fulfilled.

There are many types of uses for propensity scores (matching, stratification), and IPTW were used in this analysis. IPTW were used

instead of matching in order to avoid restrictions in the sample, as the sample size is quite large and the number of individuals experiencing a health shocks comparatively low (about 2,000 individuals). Given the common support, we expect no significant difference in estimates from those two techniques. IPTW are calculated as follow (Austin, An introduction to propensity score methods for reducing the effects of confounding in observational studies, 2011):

$$w_i = \frac{t_i}{\pi_i} + \frac{(1 - t_i)}{(1 - \pi_i)}$$

which follows the notation laid out above. These weights are then used as sampling weights in the first difference regressions.

By combining a first difference approach to IPTW, both differences on observables and heterogeneity based on time-invariant unobservable characteristics, which would otherwise bias estimates, are accounted for.

6.2.2. Data

The dataset used in this chapter is the China Family Panel Studies (or CFPS) (please see Chapter 3 on data and methods for full information). The sample of analysis consists of about 16 000 individuals aged between 16 and 55 for which data is available for both waves (sample size depends on response rate for each variable). In line with other studies on labour force participation in China, a limit is defined at 55 years of age (Liu, 2010). There is a sharp decline in labour force participation between 55 and 60 in the sample.

6.2.3. Measurement

Health shock

By contrast to the Ningxia project data, CFPS does not contain data on diagnosis or cause of hospitalisation, and alternatively, health shocks are defined as a hospital admission for seven days or more. This definition of health shock using hospitalisation data has been used in previous works (see literature review for discussion). In this data, average length of stay for all hospitalisation in the sample is 10.54 days (slightly higher in urban areas, 11.87 days compared to 9.69 days). The threshold is set at 7 days based on the observed high average length of stay and the thresholds used in the literature. The regressions are also repeated using different thresholds (between 1 day to 14 days) to test the sensitivity of the results to the threshold set (results presented and discussed in Section 6.3.4).

Labour force supply

Information on labour force supply is compiled using questions pertaining to work for different jobs in the following sectors (allowing for more than one job in each sector): agricultural, wage, and business or self-employed. The information is collected in the following manner:

- Individuals are asked how many months they have been working in the sector and/or job under consideration in the previous 12 months
- If the individual has worked more than 1 month, they are then asked to declare how many days a month they have worked (on average for each working month)

- Finally, individuals are then asked to report how many hours per day (on average for each working day).

For each category of employment, a 12-month average is compiled and all categories are added up for each individual in order to generate a variable for the total work hours per day in the last 12 months. It is not uncommon for individuals to combine several jobs, especially for individuals with an agricultural activity. Unlike other studies, work hours per day are calculated on a 7-day week (including weekends) rather than on the basis of a regular 5-day week. The number of hours per day is also capped at 15 hours (less than 0.2% of the sample is excluded by the capping of work hours).

In 2012, the variable on work hours for agricultural work was slightly changed from its previous version in 2010. Instead of declaring an average over the year, respondents were instead asked to report the number of months worked on their farms, and then to provide two averages of work hours, respectively for the busy and the off farm season. The main challenge with computing an average using this data is that respondents were not asked how many months they worked in each of these two categories. A review of the agricultural literature on the subject was conducted to identify the number of months that were usually considered busy seasons (labouring the crops, plantation and harvesting) for the main three types of crops cultivated in our sample: rice, wheat and corn. To compound matters, some farmers might be cultivating several crops at the same time, but this information was not collected in the survey.

The literature review shows that on average, busy seasons (planting and harvest) amounted to 6 to 8 months in the year depending on the crop. As a result, it was decided that the busy season would account for 8 months on average in the year. If an individual declares working for less or equal to 8 months on agricultural production, then these months are considered in the busy season. It is a reasonable assumption given that migrants, for instance, might combine agricultural work with migrant work and return for the busy season. It is unlikely that individuals with another separate economic activity would choose to work on their crops during the off-farm season. For individuals performing agricultural work for more than eight months, number of hours per day is weighted by 0.67 for the busy season and 0.33 otherwise.

In addition to work hours, a dummy variable for inactivity is defined based on whether a person reports zero hours of work.

Table 6-1 provides summary statistics of health shocks as well as the different dependent variables (hours of work and inactivity). This table shows that the means and distributions across the two years and rural and urban areas are quite stable. Comparing the hours of work in 2010 and 2012 in rural areas, the means (respectively 4.27 and 4.43) and distributions also do not raise major concerns, which is reassuring given the difference in methodology for the calculation of the average hours of work between the two waves.

Table 6-1. Summary statistics for health shock and main outcomes of interest

Variable	Mean	Standard deviation	Skewness
Health shock (under 55)	0.41	0.199	4.59
Hours of work			
Total	4.43	4.03	0.53
Rural	4.35	4.02	0.62
Urban	4.52	4.04	0.42
2010	4.36	3.88	0.44
2012	4.50	4.17	0.58
Difference between two years	0.20	4.59	0.11
Inactive (hours of work = 0)			
Total	0.28	0.45	0.93
Rural	0.26	0.44	1.06
Urban	0.31	0.46	0.78
2010	0.29	0.46	0.91
2012	0.29	0.45	0.94
Difference between 2 years	-0.06	0.51	-0.11

* Minimum and maximums are not reported in the table. Hours of work are comprised between 0 and 15 hours a day, while inactivity and health shocks are comprised between 0 and 1 (dummy variables).

6.3. Results

6.3.1. Descriptive statistics

Health shocks

In 2012, 6.33% of adults (aged 16 years of age and above) had reported a health shock (as defined in terms of length of admission) in the preceding 12 months. Table 6-2. provides a summary of the 'prevalence' of health shocks across age groups, educational attainment, wealth quintile and gender. Health shocks are common among older populations, and populations with lower educational attainment. The rate of shock is slightly lower for households reporting annual consumption below the community median, although the difference is not very large.

Table 6-2. Rate of shock across socio-economic groups

Variable	Rate of shock
Age groups	
16-24	1.91%
25-34	3.49%
35-44	4.05%
45-54	6.25%
55-64	9.24%
65+	13.44%
Consumption	
Below median	5.55%
Above median	6.76%
Education groups	
No school	9.32%
Elementary school	6.84%
Middle school	5.09%
High school or more	4.32%

Table 6-3 provides a summary of the key differences between the sample of two individuals. There are systematic differences between individuals who experienced a health shock and those who did not on a number of socio-economic characteristics.

First of all, individuals who experienced a health shock are likely to be much older than the rest of the sample, which is not surprising and also observed in many other health shocks studies (including in this manuscript). Moreover, individuals who experienced a health shock have lower education, declare lower household consumption, which is also more likely to be below the county median. With regards to their labour force participation, individuals who experienced a health shock are more likely to be out of the labour market. This is partly explained by the fact that the average age of individuals experiencing a health shock is 53 years old, significantly higher than for individuals who did not experience a shock. If we restrict to the sample of individuals age above 45, none of the differences in socio-economic characteristics are not significant.

Table 6-3. Comparison in individual and household baseline characteristics between individuals experiencing a health shock versus individuals not experiencing a health shock.

	Mean (shock=0)	Mean (shock=1)	Significance	P-value (1)
Age	44.58	53.49	***	0.00
Gender (% male)	48.79	45.71	**	0.04
No education (%)	27.40	39.32	***	0.00
Han (%)	91.54	93.99	***	0.00
Married (%)	78.88	87.13	***	0.00
Economically active (%)	63.24	55.39	***	0.00
Urban (%)	46.54	45.51		0.22
Access to water (%)	58.61	58.62		0.56
Household (HH) size	4.02	4.24	***	0.00
HH consumption	33309	30502	***	0.00
HH consu below median (%)	55.32	52.65	*	0.05

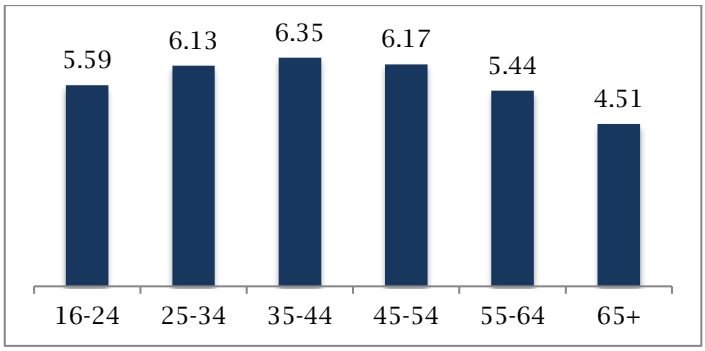
(1) Obtained from Chi2 test for categorical variables and t-test for continuous variables

*p <0.1, ** p<0.05, *** p<0.01

Labour force supply

71.2% of the working age population declares having worked in the preceding 12 months. There are important differences in labour market participation between men and women, as 63.5% of working age women report to be active, against 80.3% of men. On average, women work fewer hours than men (conditional on participation): women work 5.38 hours a day against men who declare an average of 6.43 work hours a day. The number of worked hours conditional on participation decreases also with age.

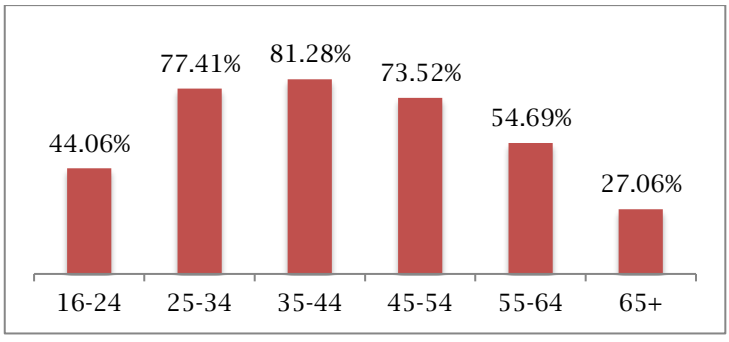
Figure 6-1. Number of work hours by age band (for both genders), baseline (2010)



NB. These figures are conditional on participation

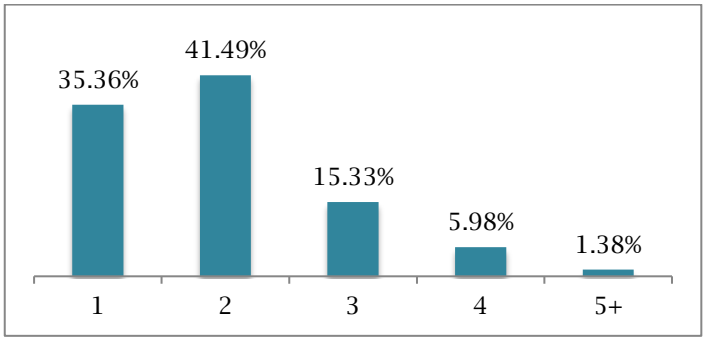
Figure 6-2 shows the average labour market participation rate by age group. Labour force participation decreases rapidly after the age of 55. Again, there are important differences between men and women, and this decrease in labour force participation after 55 is particularly important in the case of women. At baseline year (2010), only 34.2% of women aged 55 years or over had worked in the past 12 months, against 52% for their male counterparts.

Figure 6-2. Participation rates by age group (both genders), baseline (2010)



Finally, as shown on Figure 6-3 (below), 76.85% of individuals live in a household with one or two economically active individual[9] (declaring positive hours of work in the past year). On the other hand, it is worth noting that the average number of working age adults per household is 2.03; and that 73% of individuals in our sample live in a household with more than one working age individual. These statistics indicate that there are opportunities for other household members to start working in order to compensate for the potential loss of work when shocks occur.

Figure 6-3. Percentage of individuals living in households with the following number of economically active adults (for both genders), baseline (2010).



With regards to the relationship between labour force and health shocks, individuals who have experienced a shock in 2012 report lower hours of work compared to those who haven't experienced a shock at baseline (2010) (3.22 hours against 3.88 for those without a shock). This difference is partly explained by the fact that individuals reporting a shock are usually older (on average 9 years older). For instance, if we look

[9] This discussion excludes households with no economically active individuals, which represented 15% of the sample and consisted mostly of households with no observed individual (no response) or households with only elderly. 13.4% of adults in the sample are aged 65 and above. Households with no economically active individuals usually have a larger share of elderly (share of individuals aged 60+) and children in the household compared to household with at least one economically active.

at the same differences at baseline, this time conditioning for age between 40 and 55, the differences become insignificant[10] (4.66 hours for individuals who have experienced a shock versus 4.86 hours for those who have not experienced a shock, p=0.23).

6.3.2. Regression results

The results in this section are presented first for the impact of health shocks on own labour force supply, then on spousal labour force supply and finally on the work supply of other working-age household members (excluding spouse). Each of these subsections are presented as follow. The results for the entire sample, conditioning for employment, for urban and rural samples are presented in a first table. Subsequently, further differences based on gender, household head status, and age (young workers versus older workers) are discussed in a second table.

Own labour force supply

Health shocks lead to a decrease in overall hours of work, significant only in rural areas ($p<0.01$).

[10] A t-test was used to compare the means between the two populations.

Table 6-4. Summary of estimates of the impact of health shocks on own labour force supply

	(1) All	(2) Active	(3) Urban	(4) Rural
Health Shock	-0.508*** (0.127)	-0.354** (0.145)	-0.267 (0.203)	-0.724*** (0.152)
Mean of the dep. va.	4.43	6.30	4.52	4.35
Standard deviation	4.02	3.40	4.03	4.02
Observations	16,442	10,087	7,242	9,200

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock). Standard errors are clustered at the village-level.
The second column presents results for individuals who declare positive hours of work. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
NB: Health shocks are defined as a hospital admission lasting for seven days or longer.
*** p<0.01, ** p<0.05, * p<0.1

Conditional on employment in both years, a significant drop in work hours (although of a smaller magnitude than without conditioning) is also observed for the entire sample, significant at the 1% level. These results show that as well as a reduction in work hours, individuals who experienced a health shock might drop out of work altogether.

Results presented in Table 6-5 also show differential in impact based on gender, age and whether the member who is sick is a household head.

Table 6-5. Summary of estimates of the impact of health shocks on own labour force supply (decomposition)

	(1) Men	(2) Women	(3) HH head	(4) <40 years	(5) 40-55 years
Health Shock	-0.395*	-0.572***	-0.384*	-0.319	-0.602***
	(0.218)	(0.164)	(0.204)	(0.217)	(0.158)
Mean of the dep. va	5.35	3.67	5.82	4.14	4.76
Standard deviation	4.05	3.81	3.91	4.02	4.00
Observations	7,660	8,782	6,420	8,210	8,283

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock). Standard errors are clustered at the village-level. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
Health shocks are defined as a hospital admission lasting for seven days or longer.
*** p<0.01, ** p<0.05, * p<0.1
NB: these results are not conditional on employment (whole sample), include zeros

Although significant for both genders, the estimated impacts are smaller on men compared to women, for which the effects are also significant at the 1% level. The impact on work hours of women is large. Considering that the average work hours (not conditional on participation) for women is 3.67 across both years, a reduction in 0.572 hours a day corresponds to on average a reduction in 15.5% of work hours or 56.9 days of work per year.

These gender differences in labour force response have been documented in previous work, although it is unclear what these differences are attributable to. In Liu's works (2010), health shocks have a negative

impact on the work hours of older workers and men only. These results do not align with results of this analysis, which finds a strong significant impact on female labour supply. On the other hand, other similar empirical studies (in different settings) find higher reduction in labour force supply following a health shocks in women than men (Riphahn, 1999; Pohl, Neilson, & Parro, 2014). Pohl, Neilson, & Parro (2014) further postulate that lower labour market attachment for women can explain these gender differences. Labour market attachment is related to individual's proximity to the labour force and will influence individual's participation in the labour force. Whilst commonly used in the literature, there is no accepted way of quantifying and measuring this concept. Using the assumptions of the Grossman model (1972), individual's attachment to the labour market could be determined from the reservation wage (wage at which an individual will decide to participate in an economic activity), budget constraints as well as other preferences (*e.g.* for leisure or time spent on household chores and education of children). Such information is not available in the data (*e.g.* individual wages). Given the lower number of work hours recorded for women compared to men and knowledge of the cultural norms prevalent in rural China, one could postulate that women have a lower labour force attachment compared to men.

Interestingly, work hours of younger workers do not appear to be significantly affected by health shocks (and the estimated effect seems to be small and is not significant). Again, the attachment to the labour market could explain differences in response between these two groups.

In addition, the impact of health shocks on inactivity is estimated to understand whether individuals reduce their work hours or exit the labour market completely following a shock (full regression tables in Appendix B14 and B15). Again, we find a similar pattern: health shocks increase by four per cent the probability of being inactive ($p<0.05$) of women (but not significant for men). The effects are also only significant for older adults (similar pattern as the one shown above).

Spousal effect

Spousal health shocks lead to significant increase in work hours in rural areas, not significant in the case of urban areas ($p<0.1$).

Table 6-6. Summary of estimates of the impact of health shocks on spousal labour force supply

	(1) All	(2) Active	(3) Urban	(4) Rural
Health Shock	0.230 (0.153)	0.141 (0.166)	0.0497 (0.244)	0.376* (0.221)
Mean of the dep. va	4.79	6.34	4.91	4.64
Standard deviation	4.01	3.40	4.02	3.99
Observations	9,984	6,892	4,760	5,224

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for a spousal health shock). Standard errors are clustered at the village-level.
The second column presents results for individuals who declare positive hours of work. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$
NB: mean of dependent variable is different from the figures provided above as it only includes married individuals. Health shocks are defined as a hospital admission lasting for seven days or longer.

These differences between rural and urban areas might be attributed to a more constraints on urban labour markets compared to rural areas. Most rural households in the sample are engaged in farming work (80% individuals lived in a household engaged in farming and forestry), usually on a small-scale plot; and labour substitution is relatively easier in this context (compared to urban areas).

Table 6-7. Summary of estimates of the impact of health shocks on spousal labour force supply (decomposition)

	(1) Men	(2) Women	(3) HH head	(4) <40 years	(5) 40-55 years
Health Shock	0.137	0.386*	0.144	-0.162	0.402**
	(0.244)	(0.218)	(0.251)	(0.346)	(0.187)
Mean of the dep. va.	5.97	3.77	5.97	4.79	4.80
Standard deviation	3.88	3.82	3.89	4.01	4.00
Observations	4,579	5,405	4,685	4,150	5,834

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for a spousal health shock). Standard errors are clustered at the village-level. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
*** p<0.01, ** p<0.05, * p<0.1
NB: mean of dependent variable is different from the figures provided above as it only includes married individuals. Results not conditional on employment (includes zeros). Health shocks are defined as a hospital admission lasting for seven days or longer.

Again, a spousal health shock is only significant for women (although only at the 10% level). This difference in results is in line with existing literature suggesting a larger response from women than men to spousal health shock (Liu, 2011; Coile, 2004). In addition, these gender

differences can also be partly explained by the fact that working-age men tend to work more hours compared to their female counterparts: in our data, at baseline, men work on average 45% more hours not conditioning on employment, and 19% if conditioning on employment. Given this higher workload, men might find it difficult to increase their labour supply in response to a spousal health shock, compared to women. In addition, the marginal return to additional work hours might also be lower. The same rationale could apply for the differences in response between younger and older workers.

Other household members effect (Summary results in Annex 6-2)

No significant impact of health shocks on the labour supply of other household members (other than spouse) has been detected. Not only the coefficients are insignificant in the case of other household members (for all groups), they are also very small in magnitude: the point estimates obtained for other household members are about ten fold smaller than for spouses.

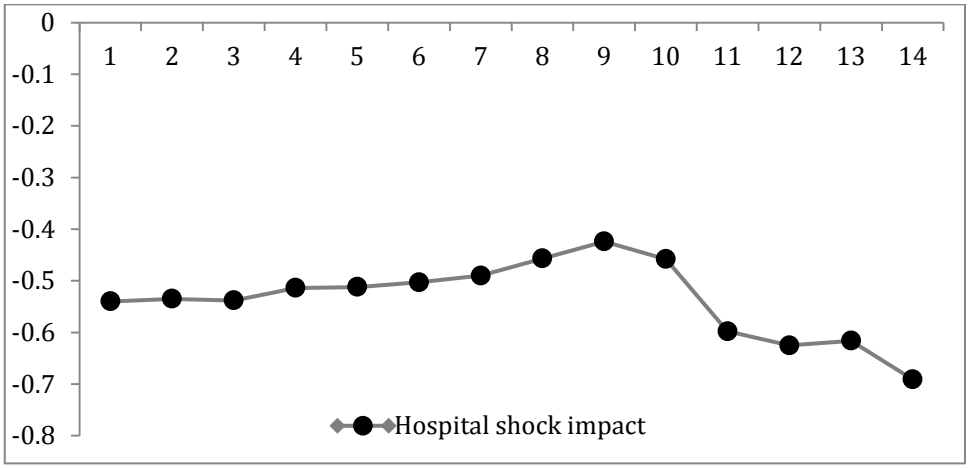
6.3.3 Sensibility to the day threshold for health shock definition

Unlike the use of diagnosis, the use of day-threshold is not theory driven and the decision of setting a 7-day threshold for the definition of the health shock was made based on similar literature. There might be a concern that results presented in this chapter might be sensitive to this 7-day threshold.

In order to test for this sensitivity of threshold, a set of health shocks is defined using day thresholds ranging from 1 to 14 (admission for one

day or more, admission for two days or more, etc.). Propensity scores are calculated again for each variable (not shown in manuscript).

Figure 6-8. Variations in estimated impact of health shocks on own work hours using different thresholds for health shock variable.



NB: The individual dots correspond to the estimated impact using the same specification as above (first difference combined with inverse probability weights used as sample weights) performed on the entire sample (age 16-55), not conditional on baseline employment.

There is no observed major change in magnitude and significance of the health shock impact on work hours from the use of any threshold of 1 to 11 days. Nonetheless, if applying a threshold of 11 day or more, we observe a much larger effect (*e.g.* of about 20% larger in magnitude for a threshold set at 14 days compared to the 7 day threshold – for which results are presented in this chapter). This suggests that the impact of work hours increases with the severity of the condition, as proxied by the length of stay.

6.4. Discussion and limitations

The analysis carried out in this chapter showed a large impact of health shocks on own work hours; largest for women, older workers and in rural areas. For instance, in the case of women, this impact of health shocks corresponds to a 15.5% reduction in work hours or 56.9 days in the year (versus a 7% reduction for men or 24.8 days of reduction per year). As hypothesized in the section on the theoretical background, the observed drops in work hours can be immediate following the shock (*e.g.* time spent in the hospital, convalescence period) or lasting effects caused by physical weaknesses or the development of a new disability. It is worth noting that other studies have put forward much larger effects (Gertler & Gruber, 2002).

These results, while not in line with previous work carried out in China, are consistent with similar studies in the literature (Coile C. C., 2004; Pohl, Neilson, & Parro, 2014). One possible explanation for these differences based on gender and age might be the lower attachment to labour force market in the case of women and older workers. Results might be only significant in rural areas due to different employment constraints (*e.g.* work hours might be enforced more systematically in urban areas), or perhaps differences in support mechanisms. There is also evidence of increase in labour force supply following a spousal shock, with similar patterns (significant in rural areas, for women and older workers). Again, this might be explained by contextual factors. Most households in rural areas work on small-scale farm production, where the transfers of labour might be facilitated compared to urban areas.

Moreover, households typically work together on one plot and there is no issue of transferability of skills. Equally, women and older individuals spend fewer hours on productive activities, and there might be more scope for them to increase their labour force supply if needed. Beyond the spousal response, results show no significant impact of health shocks on other household's labour force supply.

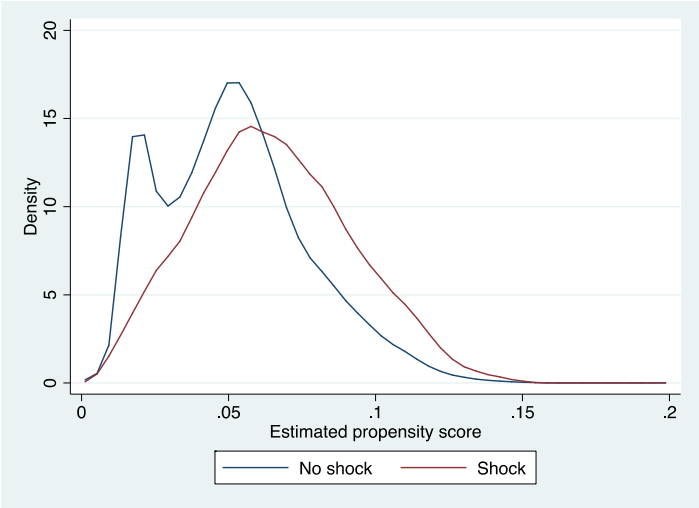
In short, these results show that individuals affected by a health shock reduce labour supply and that spouses expand their labour supply to compensate for these losses in work hours (only significant for women, older workers and in rural areas) and avoid drops in income.

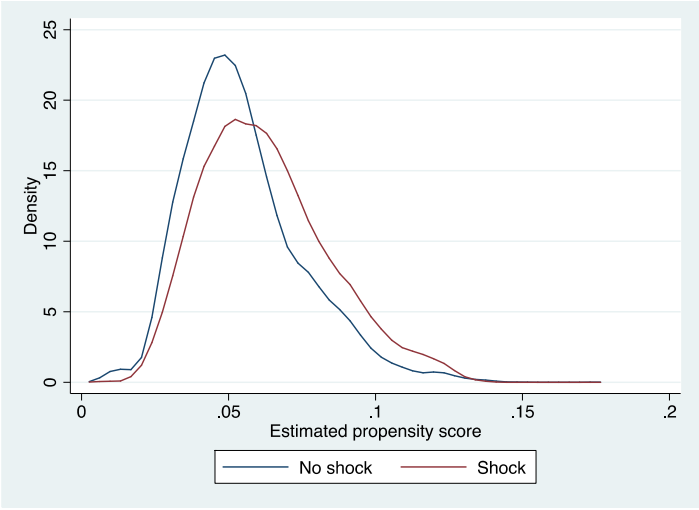
Fitting these conclusions with the results presented in Chapter 4 (showing only a decrease in income for shocks experienced by household heads), the following narrative emerges. First of all, there is evidence of significant intra-household labour reallocation in rural areas and for certain groups (*i.e.* women and elderly), which can help explaining that health shocks do not necessarily lead to a drop in income (except for shocks experienced by the household head). While these labour force transfers appear smaller in magnitude than the decrease in work hours (and therefore insufficient), they might be sufficient to prevent households from experiencing significant shortfalls in income (impacts of health shocks on income were negative but not significant –see Chapter 4). Nonetheless, health shocks experienced by household heads might impact income through other channels not captured in the raw measure of 'hours of work' (*e.g.* organization of the work within the household, productivity). Overall, combined with the results presented in Chapter 4,

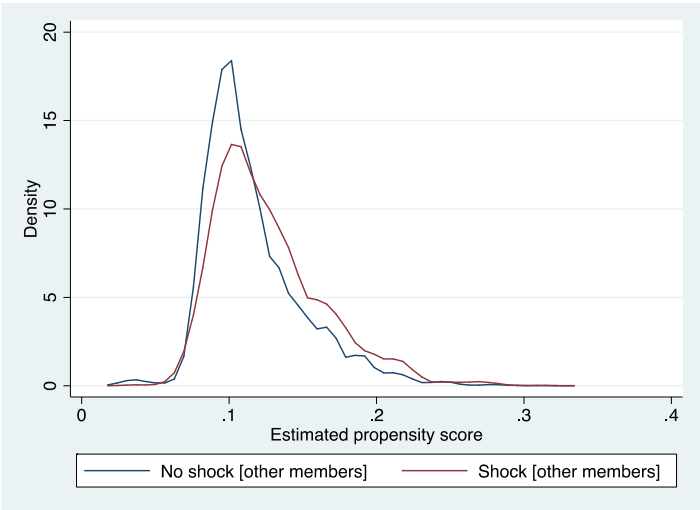
the conclusions of this chapter show significant 'private efforts' within the household to avert the negative impact of reduced work hours (from the individual who experienced a shock).

There are several limitations to this study. First of all, the impact of health shocks on non-leisure non-work time (*e.g.* household chores) are not measured in the CFPS and our analysis solely relies on work hours. Moreover, as the survey collects information on work hours that are averaged over a year (*e.g.* "how many days or hours did you spend on average working"), the variable of work hours might be subject to measurement errors (in this case recall errors). It is unclear whether these recalls errors would attenuate or accentuate the impact of shocks on work hours. However, the results found were consistent and unaffected by changes in model specification (*e.g.* logarithmic transformation). The first difference regressions were also performed without IPTW and the estimated effects were also significant but of a larger magnitude. The coefficients obtained through IPTW can be interpreted as the regression estimate over the region of common support, *i.e.* it reduces the influence of individuals who presented very different observable characteristics between the control and the health shock group (by assigning a low weight to those individuals). Such adjustment might be responsible for the observation that IPTW seems to attenuate the effects of health shocks on work hours.

Annex 6-1. Density of estimated propensity score own health and spouse (top: own health shock, second: spousal health shock, third: household health shock).







Annex 6-2. Summary results for regressions on the impact of health shocks on other household member work hours

	(1) All	(2) Active	(3) Urban	(4) Rural
Health Shock	0.055	0.009	0.149	-0.024
	(0.119)	(0.138)	(0.150)	(0.167)
Mean of the dep. va	4.51	6.23	4.59	4.43
Standard deviation	4.01	3.40	4.02	4.01
Observations	15,031	10,888	6,960	8,071

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for a spousal health shock). Standard errors are clustered at the village-level. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
*** p<0.01, ** p<0.05, * p<0.1

(continued)

	(1) Men	(2) Women	(3) HH head	(4) <40 years	(5) 40-55 years
Health Shock	0.018	0.1	-0.008	-0.014	0.091
	(0.166)	(0.138)	(0.171)	(0.175)	(0.158)
Mean of the dep. va.	5.45	3.63	5.82	4.28	4.76
Standard deviation	4.02	3.81	3.91	4.02	4
Observations	7,006	8,025	5,809	7,487	7,544

Each estimate corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for a spousal health shock). Standard errors are clustered at the village-level. All regressions are done with the following set of control: age, marital status, education, number of working age members in the household, number of children, dummy for whether the individual is a farmer, whether the household has running water, a separate toilet and electricity and average consumption at the village level.
*** p<0.01, ** p<0.05, * p<0.1
NB: results are not conditional on employment (include zeros)

Chapter 7. Accounting for health shocks in the decision to return of rural to urban migrants

It is widely acknowledged that China's internal migration flows are the largest in the world. In 2014, the number of migrants in China was 273.95 million, of which 168.2 million left their home province altogether and travelled long distances to work in cities (National Bureau of Statistics China, 2015). This increase in rural-to-urban migration is one of the most important developments in the Chinese rural economy and of central importance to the country's progress on poverty alleviation (World Bank, 2009). At the macro-level, the relaxation of the institutional residency system (*hukou system*[11]), rapid urbanization and economic growth in the cities are considered the main drivers of internal migration flows (Jialu, 2011). In 2014, the average annual disposable income in urban areas was 20167 Yuan per capita, roughly double the average annual disposable income in rural areas (10489 Yuan) (National Bureau of Statistics China, 2015). At the individual level, the determinants of migration have also

[11] China has a unique system of household registration that originates from Ancient China and was a very powerful means of population migration control until the 1980s reforms. Today the hukou does not restrict people from moving physically between locations, but it does attach individuals to a particular residency location, where all social and public services can be accessed. Transferring hukou is very difficult, and most rural to urban migrants do not do so. As a result, migrants often face problems of accommodation and employment, and are not covered by health insurance whilst in the city (the vast majority). Moreover, children of migrants cannot attend schools in the cities.

been researched thoroughly. In particular, the relationship between health and mobility has been established in the case of China; in line with the international literature on the health selectivity in migration, and the *salmon bias hypothesis*[12] (Tong & Piotrowski, 2012).

In the previous chapter, health shocks were associated with changes in work supply, significant for rural workers. However, this analysis did not cover migrant workers who worked outside of the province and it is unclear what the impact of health shocks on migrant workers more specifically is. There is a thin existing literature on the health of migrants and returnees in China from a descriptive perspective (Zhang, Liu, Zhang, & Wu, 2015). In order to bridge this gap, this chapter will investigate the relationship between health shocks and the decision to return of rural-to-urban migrant workers using the Ningxia data. More specifically, it seeks to estimate the impact of own health shock (as defined as an acute admission) on the decision to return, as well as the impact of health shocks experienced by other household members in the home communities.

This chapter is organized as follow. Section 7.1 describes the background and presents the hypotheses on health shocks and return migration under which this study is undertaken. Section 7.3 describes the empirical strategy, and Section 7.4 the results of the analysis. Finally, Section 7.4 discusses the implications of the results.

[12] This latter was first coined in the case of Hispanic migration to the United-States to refer to an observed phenomenon of return migration in the case of illness.

7.1. Review of potential channels of impact of health shocks on return decision

First of all, migration is considered temporary in this study. This is one potentially simplifying assumption, which shifts the discussion on the decision to return to a slightly different one on the timing of return. Due to the institutional constraints posed by the *hukou* system, permanent migration (with the intention to settle in the host community) only represents a marginal share of the vast population of internal migrants. Assuming migration to be temporary, the dynamic model of return migration developed by Adda, Dustmann & Mestres (2007) is a good starting point to this work. In this model, in every period *t*, a migrant updates tbeir decision to remain a migrant or to return. This decision is based on the assessment of the utility of staying for an additional year or of returning. Both utility functions are modelled using age, duration of migration, income differentials between home and host communities, savings and shocks in preferences.

In this setting, there are several channels through which health shock will discount the utility associated with migration status.

First of all, health shocks might affect the earning potential in the host community, decreasing the differentials in earnings between home and host communities. As highlighted in many studies, earnings are function of many individual characteristics, including health capital. In the particular context of China, Qin, et al. (2015) show that migrants with lower self-reported health status tend to report fewer work hours and

earnings 33% lower than healthy migrants. The importance of health capital in migrant earnings is likely to be exacerbated in the case of China, as rural migrants often face important institutional and economic constraints on the job market, and as a result, tend to mostly take on physically demanding low skill-jobs (often referred to as 'Dirty', 'Dangerous' and 'Demanding') (Qi & Niu, 2013). Démurger, Gurgand, Li, & Yue (2009) show that migrants work longer hours than their rural and urban counterparts, on average 69 hours a week. In this sense, other forms of human capital (e.g. education or experience) might not be particularly valued. Given the lack of social protection systems (and in particular labour laws), ill migrants are often sent home or face the risk of being fired.

Health shocks might also trigger shocks in locational preferences, especially if migrants have support from other household members in their home community. In addition, while health care coverage is near universal in rural areas, NCMS benefits are not transferrable to urban areas. This means that, with a few exceptions (*e.g.* own personal private insurance, employer provider insurance or enrolment to local urban scheme), migrants pay out-of-pocket for any health services accessed in urban areas (Qiu, Yang, Zhang, & Ma, 2011). Additionally, differences in relative prices of health care services in urban areas versus rural areas might lead migrants to return to seek immediate medical care or if they anticipate the need for care in the long-run. This effect could also be transmitted through the depletion of savings, thereby reducing the accumulation of financial capital and benefits of migration. This channel

is likely to be extremely important, as out-of-pocket payments for health services accessed outside of their residency area can be significant.

In this work, the impact of health shocks experienced by other household members (spouse, children and parents) on the decision to return is also explored. For this research question, the main channel through which health shocks within the household affect the decision to remain in the city would be changes in locational preferences. Lei, et al. (2011) showed that as a result of the one child policy and long distance migration, elderly are more and more likely to live alone or with an ageing spouse. As a result, there are few potential caregivers for a sick elderly; which might precipitate a decision to return for migrants. On the other hand, a health shock experienced in the household might increase the likelihood of remaining in the cities for migrants, in order to earn more money in order to send higher remittances to cover the medical costs.

7.2. Methods and measurement

7.2.1. Methods

The impact of health shocks on the decision to return is explored in this chapter. This includes health shocks experienced by migrants themselves and health shocks experienced by other household members back in the home communities. Using panel data, the following relationship is therefore observed:

$$RM_{it} = \propto HS_{it-1} + \beta X_{it} + \gamma_i + \varepsilon_i$$

where $RM_{it}$is the probability to return of individual i and time t, HS_{it-1} is a dummy variable for whether the individual has experienced a health shock (considering health shocks experienced while the individual was a migrant), X_{it} is a set of time variant covariates, γ_i individual fixed effects and ε_i is the error term.

Subsequently, the impact of health shocks experienced 'at home' (by other household members) on the decision to return is investigated using the following specification:

$$RM_{it} = \propto HS_{ht-1} + \beta X_{it} + \gamma_i + \varepsilon_i$$

which follows the same notation as above, but in this case, we are interested in the term HS_{ht}, which refers to whether a health shock has been experienced by other household members.

These two analyses are only performed on adults declaring to have migrated in previous wave. Two waves of data (2011 and 2012) are used to estimate this relationship using individual fixed effects to account for unobserved time invariant characteristics. Estimates are further broken down by gender, age group, type of migration (short-term versus long-term migrants), and severity of the health shock (as proxied by the length of admission). Further group analysis could not be performed (*e.g.* for rural-to-rural migrants or by insurance status) as no other information on the migration process was collected in those two waves. Health shocks are included as exogenous regressors. In addition, the following set of controls is used to control for individual and household characteristics: age, whether an individual is a senior, marital status, education, main occupation, dependency ratio in the household, household size, asset

quintile. The mean consumption in the village is also included, in order to account for changes in local economy that might influence migration decisions (*e.g.* wage differentials). We expect that household assets, age, dummy variable for whether the individual is a senior, marital status will have a positive effect on return decisions. On the other hand, it is unclear what the relationship between education, main occupation, dependency ratio and household size is in this context.

7.2.2. Data

The dataset used for this chapter is the Ningxia project data (see Chapter 3 for full description). The data was collected in 2009, 2011, 2012 and 2015. The household roster collects information on individuals registered in the household or considered household members. If the respondents were not present at the time of the interview, the household head (or closest proxy) would report the answers for the missing individual. In addition, the survey was collected just before and after the Spring Festival, when migrants typically return to their families[13]. This strategy ensures that information on migrants (even if they were still migrating) is collected on the same level as non-migrant residents.

In 2009, 2011 and 2012, the questionnaire collected limited information on the migration experience of individuals. In those waves, individuals

[13] The return of migrants during the Spring Festival has been the subject of extensive news coverage in the past years. An article of CNBC dated from the February 3rd 2016 refers to it as a "giant annual human migration" and mentions an estimated 2.9 billion trips registered in just under six weeks in 2016.
For more information, please see: http://www.cnbc.com/2016/02/03/spring-festival-travel-a-one-way-journey-for-many-migrant-workers.html

were asked whether they had migrated and if so, how many months they had spent working as a migrant. In order to provide more information on the migration process, a set of questions on migration were added in 2015. These included, for instance, information on the destination of migrants, occupation in the city, type of insurance, or whether they had migrated alone.

7.2.3. Problem with the data in 2015

The probability of being observed for a given year does not vary based on migration status in previous wave. Nonetheless, there is one exception: in 2015, it appears that migration in the previous wave significantly affects the probability of being observed in 2015, a relationship not observed for 2011 and 2012. For 2015, the follow-up of migrants in particular is problematic. Migration rates throughout the first three waves are stable (ranging around 26% of the total working age population), but the migration rate in 2015 is abnormally low. This drop in migration rate does not seem to be explained by an age cohort effect and is not documented in the literature. To compound matters, the return rate in 2015 also appears to be extremely high.

In order to limit the potential biases arising from inconsistent measurement, the 2015 wave is excluded from the analysis. Nonetheless, this year of data will be used to discuss the characteristics of the migration process, as this information was only collected in this wave.

The full sample contained 89,853 observations over three years of measurement. Almost 70% of households participated in all three years of survey, and 85% participated in at least two waves. Given the use of panel methods, only individuals observed for all three waves or at least for two consecutive waves are considered. Adults aged above 55 are excluded from the analysis, in order to only consider working age adults. The final sample consists of 47987 adult observations over three years.

7.2.4. Measurement

Return migration

A migrant worker is defined as an individual who has worked away from home for at least three months in the preceding year. Heterogeneity in response based on the number of migration months are explored. Following this definition, 24.8% of the working age adult sample can be designated as a migrant worker in every year considered.

Return migration is defined as follow: if an individual has declared migrating in a given wave and declared not migrating in the following wave, then he/she is considered a return migrant. Only the change in migrating status in two consecutive waves is used to define return migration; meaning that even if an individual re-migrates in a later wave, they are still counted as return. This is common to many studies on return migration, as there is no method for identifying permanent or temporary return. This definition is followed by many existing studies on return migration, in China and worldwide.

Health shock variable

Health shocks are defined as a hospitalisation for causes that are acute, likely to be sudden (and therefore unplanned in the timing of its occurrence) and to long-term impairments (García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, 2013; Gupta, Kleinjans, & Larsenc, 2015; McClellan, 1998). These include heart attacks, strokes, onset of a new cancer (except for skin cancer), diseases of the digestive system, and accidents.

7.3. Results

7.3.1. Descriptive information

Migration

Across all three years, 70.4% of adults have not been migrating in any year. For those who migrated, 44.1% of individuals migrated for one year, 33.4% for two years and 22.5% for all three years. The average months of migration is 8.4 months. 41.2% of migrants migrate for more than nine months in the year. Migrants are on average younger than rural non-migrants (30.8 years of age versus 36.9). The typical migrant profile is a young adult male (only 20.4% of migrants are female), with higher educational attainment (only 10.5% never attended school versus 24.3% in the non-migrant sample), and a slightly higher probability of being single (30.7% of migrants are single, against 27.4%).

The additional information collected in 2015 on migration is used to describe the migration process. Please bear in mind that the interpretation of this information might be subject to a bias arising from a potential change in data collection, which might have altered the quality of the data with respect to migrant workers. In 2015, 52.5% of migrants migrated within the province (to an urban area outside of county) and 34.7% migrated to a city in another province (the remaining share migrating to another rural area). In terms of industry, migrants predominately worked in the construction sector (49.1%), followed by the service industry (21.6%) and the manufacture industry (8.8%). Half of the individuals migrated alone, while 19.7% migrated with other villagers and 31.2% migrated with other family members. 9.78% declared having health insurance in urban areas. To the question "are you a return migrant?", 7% of total rural sample answered yes, and the three most common cited reasons for returning were: looking after a family member (39.2%), being too old (26.1%), being ill (13.5%).

When looking at return migrants (change from declaring to have migrated to not having migrated in two consecutive waves of data collection), 5.4% of the total adult sample can be characterized as a return migrant. Return migrants seem to be on average older than active migrants, have lower education and be married. Return migrants also seem to return to households with higher assets, consumption (but slightly lower income). The rate of return is also highest amongst 'short term migrants' (those who migrated for less than six months) (43.9%) compared to the rest of the migrant sample (37.1%). Table 7-1 provides the summary characteristics of return migrants versus active migrants.

Table 7-1. Characteristics of return migrants versus active migrants (year before return)

	Mean (return=0)	Mean (return=1)	t-test/ Chi2
Age	29.6	33.6	0.00
Gender (% male)	89.3	74.1	0.00
No education (%)	7.94	14.56	0.00
Married (%)	72.83	82.47	0.00
Household (HH) size	4.89	4.88	0.85
Average months of migration (t-1)	7.80	7.25	0.00
Asset index	-0.83	-0.76	0.00
HH consumption	23103	25301	0.00
HH income	20980	19492	0.00

Health shocks and migration

On average, 3.39% of working age adults experienced a health shock. Decomposing by migration group, respectively 2.3%, 4.5% and 4.7% of active migrants, return migrants and rural non-migrant dwellers have experienced such a shock. Exactly half of the individuals who experience a health shock during their migration time declare to return in the following year. Those who return with a health shock compared to those who remain in cities tend to be slightly older, more likely to be women and have lower education.

7.3.2. Regression estimates

The following regression estimates will be presented in the following manner. First of all, results are presented for the entire population, further decomposed by gender, whether the migrant is the household head and by age group. In a second table, heterogeneity will be explored based on previous migration status (long term versus short term) and

severity of the health shock. Full regression tables are available in Appendix B20-B22.

Health shocks are statistically significant and positively associated with return decisions for men, and a health shock corresponds to a 34% increase in the probability to return (holding all the other covariates constant). It is worth noting that the coefficient is also positive for women, but not precisely estimated. This is likely to be due to the fact that fewer female migrants are observed in the data as most migrants are men. This effect is only significant for hospital admissions for acute reasons for more than seven days. It suggests that the severity of the health shock also plays an important role in defining the decision to return. This is consistent with our hypothesis, as long hospital admission will lead to important medical costs, and therefore increase the likelihood of seeking such care in the hometown. Moreover, the length of the admission might indicate that the condition is severe and could lead to a prolonged impact on individual health beyond the hospitalization only.

Table 7-2. Return migration and own health shocks (by individual characteristics)

	(1) All migrants	(2) Male	(3) Female	(4) HH head	(5) age<40	(6) age>=40
Health shock	0.334**	0.304**	0.068	0.221	0.199	0.318
	(0.142)	(0.128)	(0.361)	(0.184)	(0.187)	(0.239)
Number of observations	3,750	2,844	906	1,802	2,328	1,334

*** p<0.01, ** p<0.05, * p<0.1. Point estimates obtained by fixed effects logit models (individual fixed effects). The dependent variable is 'return migration' and the predictor of interest is health shocks. The results are presented for all migrants, and split by length of migration and severity of health shock as proxied by length of stay (by restricting the samples). A full set of controls is included: age, age squared, marital status, educational attainment household size, number of elderly, number of children, house quality index, mean consumption in the village.
Robust standard errors are used in this analysis.
Full results are presented in Appendix B20.

Table 7-3. Return migration and own health shocks (by individual characteristics) (continued)

	(7) <9 months migration	(8) >=9 months migration	(9) admission < or = 7 days	(10) admission over 7 days
Health shock	0.336	0.221	0.284	0.393**
	(0.220)	(0.310)	(0.189)	(0.176)
Number of observations	2,572	1,430	3,750	3,750

*** p<0.01, ** p<0.05, * p<0.1. Point estimates obtained by fixed effects logit models (individual fixed effects). The dependent variable is 'return migration' and the predictor of interest is health shocks. The results are presented for all migrants, and split by length of migration and severity of health shock as proxied by length of stay (by restricting the samples). A full set of controls is included: age, age squared, marital status, educational attainment household size, number of elderly, number of children, house quality index, mean consumption in the village.
Robust standard errors are used in this analysis.
Full results are presented in Appendix B21.

Health shocks experienced by other household members do not seem to have a significant impact on return migration (for any of the population under consideration). It is worth noting that the estimated effects seem to be small in magnitude, in comparison with the effects displayed above. Those results are in line with those presented in Chapter 4, which show an increase in remittances following the occurrence of a shock in households with a migrant worker. Combine, they show that households rely on the migration process (and its remittances) when faced with a health shock. This coping channel was put forward in the case of Mexico by Ambrosius & Cuecuecha (2013). In the case of China, only anecdotal evidence can be found in Ping & Zhang (2008).

Table 7-4. Return migration and household health shocks (by individual characteristics)

	(1) All migrants	(2) Male	(3) Female	(4) HH head	(5) age<40	(6) Age>=40
Health shock	0.133 (0.104)	0.153 (0.120)	0.0461 (0.217)	0.121 (0.158)	0.0872 (0.129)	0.281 (0.190)
Nb of obs.	3,536	2,678	858	1,662	2,220	1,232

*** p<0.01, ** p<0.05, * p<0.1. Point estimates obtained by fixed effects logit models (individual fixed effects). The dependent variable is 'return migration' and the predictor of interest is health shocks experienced in the household. The results are presented for all migrants, and split by length of migration and severity of health shock as proxied by length of stay (by restricting the samples). A full set of controls is included: age, age squared, marital status, educational attainment household size, number of elderly, number of children, house quality index, mean consumption in the village. Robust standard errors are used in this analysis.
Full results are presented in Appendix B22.

7.4. Conclusions and limitations

The rate of out-migration has increased steadily in the last two decades in China, yet little is known on the process of return and its determinants. This chapter takes advantage of a large-scale panel dataset on Chinese household in the province of Ningxia, which has put particular attention to collecting information on migrants, through the data collection strategy.

The analyses showed a statistically significant positive impact of health shocks on the decision to return for men, but not for women (although the coefficient is not precisely estimated). Additionally, the severity of the health shock (as proxied by the length of the hospital admission) has an impact on the migrant's decision to return as well. On the other hand, there was no significant impact of a health shock experienced in the household (in the home community) on a migrant's decision to return. Although insignificant, this analysis shows that the impact of a household health shock would be of a relatively small magnitude and negative.

Return migrants have often been presented as actors of change and development in rural China (Démurger & Xu, 2011), through accumulated financial, human and social capital. Return migrants with a health shock might be ill equipped to similarly contribute to rural development. A separate analysis (Annex 7-1) shows that the proportion of return migrants with a health shock returning to farm work after three or four

years of return is higher than for return migrants without a health shock. In addition, return migrants with a health shock are less likely to re-migrate and declare higher health expenditure in subsequent waves. This hints that these migrants might not correspond to the picture of success as depicted in Démurger & Xu (2011), and that further research on the conditions under which returnees come back is important to understand the link between return and rural development.

This study presents several limits. First of all, we were unable to account for factors linked to the migration process itself (only controls in home communities), as our dataset did not collect such information. For instance, the destination of the migrant and migration history is not known. These are two of the main predictors put forward in the literature on return migration. Migrants with health shocks might exhibit unobserved characteristics that can be also linked with their migration intentions (in terms of migration duration, for instance) or migration experience. Information on the destination of migration could have further informed and supported this approach but was unavailable in the data collection. Consequently, the comparison of the long-term outcomes of return migrants needs to be considered with caution. Finally, the last limitation refers to the fact that information on the location of the hospital admission was not recorded properly in the data. As a result, it was not possible to understand fully the timing of events, *i.e.* whether the migrant returned to seek care or whether he sought care at destination following the health shock and subsequently returned. More research on the care seeking behaviour of migrants following a health shock should be undertaken.

Despite the above stated limitations, this study highlights some of the pervasive and complex effects of health shocks on households, which are rarely considered in health shock studies. This study adds a new outcome for consideration in the body of literature of health shocks. In the context of China, this effect on return migration and on returnees' long-term outcomes should be carefully considered. Given the sheer size of the migration process in China and this evidence on the link between return and health shocks, particular attention to the health of migrants in order to reap the benefits of migration for both the host and the home communities.

Annex 7-1. Differences in long-term outcomes between return migrants with a health shock and return migrants without a health shock.

The long-term socio-economic outcomes of return migrants with a health shocks versus those who return without a health shock are compared using the 2015 wave. The objective is to explore whether the literature on successful return migrants accounts for the growing share of return migrants with a health problem. Since return migration was shown to positively impact rural development (especially through human and social capital brought back to the home areas), it would be of interest to asses whether this effect is observed for return migrants with a health shock. This argument was first put forward in a qualitative study of ill return migrants that shows that return migrants with illnesses pose an important burden on households (Chen, Ding, Cook, & Pong, 2014).

For this analysis, households with a return migrant in 2011 and 2012 are selected. From this group, two separate groups are defined based on whether the return migrant has experienced a health shock or not. These results are limited by the fact that there is a larger rate of attrition between 2012 and 2015 and the number of return migrants followed-up across all three waves is small. This is a simple, very descriptive comparison of outcomes, that is not standardized. The use of propensity scores could be explored, but with the exception of age and ethnicity, there are no significant pre-shock differences between the two groups (and even so, these differences are also quite small).

Difference in means in long-term socio-economic outcomes (2015) of return migrants, by treatment (health shock) group

	Treated (shock=1)	Control (shock=0)	Difference	Ttest /Chi2
Poor consumption*	36.18	37.17	0.08	0.83
Migration	13.89	24.00	12.11	0.04
Labourer	20.35	29.75	9.40	0.03
Farmer	61.70	73.45	11.75	0.01
Household income	29974	33787	3813	0.23
Health expenditure	6419	4587	-1832	0.07
Outstanding loans	33707	31632	-2075	0.67
Asset index	0.04	0.14	0.14	0.27
Household consumption	60908	55405	5503	0.40

* food consumption below the median in their county

As shown on the table above, three significant differences between the return migrants with a health shock and those without a health shock are observed. First of all, the return migrants with a health shock consistently report higher household medical expenditure compared to healthy returnees ($p<0.1$), suggesting that the health shocks have long-lasting effects on medical expenditure. In addition, return migrants with a health shock are more likely to return to farming and less likely to become laborers. Remigration rates in 2015 of return migrants with a health shocks are also lower ($p<0.05$).

Chapter 8. Conclusion and discussion

The aim of the work undertaken in this manuscript is to estimate household response to health shocks in a comprehensive manner, using multiple outcomes and indicators of household well-being and socio-economic status, thus going beyond the sole focus on medical expenditure. From a general perspective, a first observation can be made about how intricate the household response to health shocks is: the estimates presented in this manuscript have shown statistically significant impacts on outcomes ranging from household medical expenditure, food consumption, to spousal labour force supply or return migration. Another general comment is that the timing of these effects is also complex and needs to be further investigated in future studies. Only one study has sought to identify the long-term effects of shocks (see Chapter 2 for discussion) (García-Gómez, Van Kippersluis, O'Donnell, & Van Doorslaer, 2013). The evidence presented in this thesis shows that the long-term effects are likely to be very large and significant and highlights plausible channels of impoverishment due to health shock through inefficient coping strategies.

8.1. Summary evidence of the results

Using the three original questions presented in the introduction of the thesis to structure this section, we provide a succinct summary of the evidence presented in this manuscript.

8.1.1. What are the total immediate economic costs associated with health shocks?

Unsurprisingly, Chapter 4 shows that health shocks lead to statistically significant increases in medical expenditure ranging between 1,607 Yuan (for a woman) and 2,624 Yuan (for households experiencing multiple shocks). This is equivalent to an increase of 67% to 109% of household yearly medical expenditure. In addition to those costs, health shocks lead to a significant reduction in income for shocks experienced by the household head, ranging around 1,000 Yuan or 5% of the yearly total household income. This impact on household income is not significant for shocks experienced by other household members but the household head.

In Chapter 6, we have seen that this decrease in income can be partly explained by a reduction of labour force supply (labour force participation and work hours) following a shock. This impact is significant in rural areas, for all working age adults (although only at the 10% for men and household heads). Estimated effects are strongest for women and older workers. The estimated impact of health shocks on work hours is equivalent to a 15.5% reduction in annual work hours for women. Further, health shocks lead to a 4% increase in the probability to

exit the labour force market completely for women. In the same vein, health shocks are also associated with a higher probability to return for urban-to-rural migrant workers. Rural-to-urban migrants are an important group in rural areas, as 29.4% of individuals followed in the Ningxia data have migrated at least once in any of the four waves of data collection (2009, 2011, 2012 or 2015). Moreover, migrant workers increase household income by sending remittances, which represent on average 30% of total household income in the Ningxia project data (for households with at least one migrant worker). In this setting, returns of migrant workers prompted by the occurrence of a health shock can have a large negative impact on household income.

Nonetheless, an analysis of changes in work hours within the household (in Chapter 6) showed evidence of important 'private efforts' in the household to compensate for those decreases in labour force supply from the sick individual. Health shocks lead to statistically significant increases in labour force supply from spouses only (not significant for other household members). Women and older workers, again, are more likely to increase their labour force supply following a spousal health shock. From this analysis carried out in Chapter 6, it is hard to assess whether those changes in spousal work hours are large enough to avoid a decrease in household income. However, combined with estimates obtained in Chapter 4, these results might help explaining the insignificant impact of health shocks on income when experienced by other household members but the household head.

8.1.2. What types of strategies are employed by households to cope with shocks?

As stated above, spouses (women and elderly) increase their labour force supply to avoid shortfalls in income following the occurrence of a shock. Other identified coping strategies include increases in remittances (only significant at the 10% level), uptake of debts and sale of asset (only significant for health shocks experienced by household heads). In addition, health shocks lead to a reduction in same year savings but it is unclear whether dissaving is used as a coping strategy (as the Ningxia Project Data does not collect such information).

In general, uptake of debts stands out as the single most important coping strategy for households. Health shocks are associated with an increase in debts ranging between 2,100 to 4,476 Yuan depending on the type of shock. From a qualitative point of view, roughly a third of households covered by the Ningxia Project survey declare having taken a loan in the past 12 months to 'pay for medical care'. Medical bills ranked as the most cited reason, followed by 'building or repairing of the house' and 'paying for wedding or funeral expenses'. It is worth noting that the rate of indebtedness is also quite high in this sample as 65.7% of households declare to have an outstanding loan.

8.1.3. What are some of the long-term consequences of shocks on households?

Health shocks do not seem to have an impact on same year consumption patterns; with a few exceptions. First of all, health shocks experienced by household heads lead to a decrease in food consumption ranging around 4%. Health shocks also have a positive impact on NFNM consumption, mainly driven by an increase in gift expenditure (not fully explained). This regression type is similar to those used in the existing research literature that intends to test the completeness of informal insurance network (full insurance hypothesis) or to understand the welfare consequences of health shocks.

Nonetheless, a very different pattern emerges when adopting a larger timeframe and looking at the effects of shocks not on same year consumption, but on consumption two, three or six years after the shock occurs. These results are presented in Chapter 5 and show large negative impacts of shocks, suggesting the existence of delayed costs. The effects are very large on NFNM consumption (ranging around 34%) and are statistically significant, even after accounting for other changes in household demographic and socio-economic characteristics. Interestingly, health shocks lead to a large increase in medical expenditure in the same year and two years after the occurrence of the shock (although of a smaller magnitude). However, these effects are not significant three and six years after the shock, suggesting that individuals do not experience long-term increases in medical expenditure beyond their recovery period.

Combined with the analysis on coping strategies, these results concur with the conclusions of Chetty & Looney (2006), which highlight the negative impact of uptake of inefficient coping strategies by households.

Chetty & Looney (2006) show that households might be able to smooth consumption following an idiosyncratic shock because they are risk averse (especially poor households) and will employ coping strategies to avoid a reduction of consumption levels.

8.1.4. Contributions and comparison with existing literature

As previously pointed out, there are three existing studies of health shocks in China. Several points of convergence exist when comparing the results from this present study to this existing literature. First of all, as with Lindelow & Wagstaff (2005), we find a significant increase in medical expenditure, although the estimated effect is a lot larger in this study (between 57% and 109% compared to 9% in their study). In addition, this present study shows a significant impact of health shocks experienced by household heads on income. This time, the estimated effects in Chapter 4 are smaller than those presented in Lindelow & Wagstaff (2005), which estimated impacts between 6.2% and 12.4% depending on the magnitude of the shock.

As with Liu (2010) and Lindelow & Wagstaff (2005), our results point to a negative and significant impact of health shocks on the number of hours worked. Nonetheless, Lindelow & Wagstaff (2005) did not find any significant impact of health shocks on other household members' health shocks (including spouses). Liu (2010) finds a significant impact on time use of spouses for women, which is consistent with our findings. Finally, our results on household consumption are also in agreement with the recent works from Liu (2016) (using the CHNS), which shows that

households are mostly able to cope with shocks without experiencing large decreases in consumption. In this manuscript, only health shocks experienced by household heads lead to a statistically significant reduction in food consumption. This reduction is also of a relatively small magnitude (4% of total household food consumption). Liu (2016) finds no significant impact of health shocks on any type of household consumption.

Compared to those studies, the contribution of this manuscript lies in several elements. First of all, the emphasis on the methodology of the construction of the health shock variable and understanding of the pre-shock conditions are an essential part of this work. The health shock variable used in this manuscript is consistent with a wider health economics literature (see discussion in Chapter 3) and puts a lot of efforts to support the conditional exogeneity assumption. In addition, when judged feasible and suitable, propensity score methods were applied to further control for observed confounders. These additional steps helped provide justification for the validity of this health shock approach.

Unlike any of the existing studies cited above, this manuscript discusses the household response to health shocks using a wide range of indicators of socio-economic status and coping strategies; and tries to identify potential sources of heterogeneity (based on the characteristics of the health shock and the individual who experiences the shock, as well as the socio-economic status of the household). This approach allowed us to tie in results from different analyses in order to identify channels of impacts

in one single research setting (China, more specifically rural areas). Only a handful of studies have adopted such a broad and comprehensive approach. Notably, more recent studies employ a wider range of outcomes, whereas earlier studies focus on consumption patterns, income or medical expenditure (Alam & Mahal, 2014). For instance, the recent works of Sparrow and colleagues look at the impact of health shocks on consumption, income, medical expenditure, dissaving, and borrowing in Indonesia (Sparrow, et al, 2014). Mitra, Palmer, Mont & Groce (2016) investigate household response to health shocks in Vietnam, with a specific focus on coping strategies.

Finally, another contribution is the use of a longer timeframe to understand the impact of health shocks in the long run. This additional analysis has helped illustrate the potential detrimental impact of inefficient coping strategies, and complements the theoretical literature put forward by Chetty & Looney (2006).

8.2 Policy implications

Using the Ningxia Project data, we have seen that health shocks are fairly common as 5.2% of adults experience a shock in a given wave; and that amongst individuals followed for all four waves, 13.2% of individuals have experienced at least one health shock in any of the waves of data collection. This is in line with a development economics literature on risks that ranks health problems (in more generic terms rather than health shocks) as the main adverse event to households in LMICs (Wagstaff & Lindelow, 2014). In particular, this research is very relevant to

the context of China, which has one of the most rapidly ageing population in the world with a growing burden of NCDs (Lutz, Sanderson, & Scherbov, 2008). Concurrently, China's track history of economic growth has been monumental and its recent efforts to expand and reform health care coverage commendable. The results presented in this thesis show that health shocks are common and represent a significant burden to households. As a result, they might undermine the welfare gains from expanded health insurance coverage and economic growth for households.

Several lessons can be drawn from the work contained in this manuscript. First of all, the attention paid to health expenditure in the research literature when estimating the costs associated with health shock seems to be reasonable. This research points to important increases in health expenditure following a shock, in a setting where health insurance coverage is near universal (95% in the Ningxia Project provinces). This is in line with the literature evaluating the financial protection offered by NCMS, which points to shallow and poorly structured reimbursement rates (Sun, Jackson, Carmichael, & Sleigh, 2009).

Households also have to cope with income losses, but those are partly offset by an increase in work hours from spouses, although this result is only significant in rural areas. This might be explained by the fact that households in rural areas are more likely to work on small-scale farming (where a family works on one or several plots) and that transfers of labour from the sick member to spouse is easier in this setting. While significant, it is unclear whether the increase in work hours from spouses

is sufficient to avoid a shortfall in income. More work on the household labour response and the incurred income losses should be undertaken to address those points. In particular, as shown in Chapter 7 (for rural-to-urban migrants), health shocks might lead to more complex labour force transitions. The literature in developed countries uses several types of outcomes ranging from premature retirement, change to lower-job prospects, switch to part-time jobs, and unemployment. While those types of outcomes might not be entirely relevant in the case of China (or any other LMICs), they show that the negative effects of health shocks on labour force supply go beyond just working hours or participation. Investigating how health shocks might shape labour force decisions is worthy of further research and will have important policy implications in the context of LMICs.

Finally, households seem to be able to maintain consumption in the year of the health shock. Nonetheless, Chapter 5 shows large and significant negative impacts on consumption levels in subsequent years, confirming the existence of delayed costs perhaps associated with coping strategies. As shown in Flores, et al. (2008), the study of coping strategies is key to understanding the impact of health problems on households.

Households are not always portrayed as active units of decision-making in the health economics literature. Nonetheless, this manuscript provides evidence that health shocks lead to a fairly complex household response, ranging from debts, to use of savings, increase in labour force supply to remittances. Those coping strategies are not equivalent, in terms of their contemporaneous and future impact on household welfare. Some of

those strategies are intelligent, and others generate additional costs resulting in an additional burden for households. The study of these coping strategies is crucial to the design of smart policies that can support and improve household resilience to health shocks. In particular, reducing the negative impact of coping strategies that have negative impacts on long-term household welfare could be an important policy. For instance, given the prevalent use of loans for medical reasons, it would be beneficial to households to provide opportunities for safe and affordable credit. Sound financial markets in developing countries, especially in rural areas, would also have far reaching benefits on growth and risk coping (Conning & Udry, 2007; Karlan & Zinman, 2009).

8.3 Limitations

While a lot of attention was put on the construction of the health shock variable, one of the main limitations of this study is that individuals with a shock seem to present different characteristics compared to individuals who did not experience a shock. Such differences are not problematic as long as they are controlled for. The inclusion of an extensive set of controls, the use of individual or household fixed effects, and in some cases, the use of IPTW (based on propensity scores) help accounting for those differences in individual characteristics. Nonetheless, a separate analysis of the pre-shock trend for individuals who experienced a shock in 2015 (using the 2009, 2011 and 2012 waves) showed that even controlling for differences in socio-economic characteristics, statistically significant differences in self-rated health, chronic disease prevalence and health expenditure remained (see Annex 3-2). This might be

explained by inclusion of some diagnoses included in the construction of the health shock variable. For instance, while the timing of a stroke is not exactly anticipated, individuals might experience a decrease in health status prior to the onset of such a condition. The main risk factor for stroke is hypertension, which is asymptomatic itself, but is closely related to common chronic conditions (*e.g.* diabetes), smoking, lack of physical exercise, overweight, etc. In this sense, while health shocks might be unanticipated in their timing, pre-shock health problems seem to appear in the years preceding the shock. No comparison or discussion of pre-shock health status was found in any of the reviewed studies, even for studies for which long panels and multiple measures of health status were available (*e.g.* CHNS). It is not certain that the current approach adopted in this manuscript successfully takes into account such pre-shock conditions. If not properly accounted for, the estimates presented in this manuscript might suffer from a downward bias.

Narrowing down the definition of health shocks by only looking at injuries and accidents might help accounting for such problems. Such a methodology has been considered robust in the literature (Pohl, Nielson & Sparrow, 2014), although the estimated results appear to be limited in their generalizability. In general, there seems to be a trade-off between the internal consistency and robustness of the approach, and the generalizability of the results. The definition adopted in this manuscript originally intended to strike a balance between those two elements, by adopting a validated definition of hospital admission for external causes. In China, such admissions are common and will be likely to increase given the documented rise in NCDs (Bloom, et al., 2013). Moreover, some

studies use health shocks constructed as changes self-reported health or mobility measures and do not provide much information on the underlying causes of such changes. This can be problematic when assessing the robustness of the estimates.

This study also suffers from prevalent problems in terms of reporting. All chapters rely on household surveys and self-reporting of income, consumption, loans, assets etc. The measurement and reporting of such outcomes have been shown to be problematic. For instance, looking at income reporting, Moore & Welniak (2000) show that reporting might be influenced by the interpretation of the question (including of the definition of the notion of income) and individuals' motivations in responding to the survey. Those reporting issues are sometimes accounted for in other studies by relying on a secondary source of data (*e.g.* administrative or tax data). Nonetheless, such data is seldom available in LMICs, and even then, they might present errors and difficulties in matching with household survey data. While this is a caveat to keep in mind, there are very few ways to account for these problems. Every analysis also presented its own challenges (attrition, change in indicator definition across waves, lack of proper controls etc.), which were discussed individually in each chapter.

8.4. Concluding remarks

This present study is unique in the research literature, given the comprehensiveness of the range of outcomes considered, the in-depth investigation of heterogeneous responses depending on individual or

household characteristics, and the different time frames adopted in the analysis of consumption patterns. By adopting such a research strategy, this work recognises the interdependence of household choices and their consequences on their welfare; therefore making the focus on single outcomes in the literature less relevant. From a policy perspective, this thesis also shows that attention needs to be paid in other areas (*e.g.* financial markets), in addition to the tremendous efforts made in the health sector, in order to reduce the short-term and long-term costs associated with health shocks.

Finally, this work emphasizes the need to expand our understanding of health shocks, by conducting more research on its long-term impacts and looking more closely at other variables of socio-economic status and coping that were not available to us. For instance, the impact of parental health shocks on childhood outcomes (nutrition, education investment, labour, cognitive development, human capital accumulation) could be very relevant to LMICs (including China), and document potential inter-generational effects of health shocks. Other relevant outcomes could include dis-saving, mental or physical health deterioration, and informal care of elderly.

Bibliography

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics, 112* (1), 3-56.

Adda, J., Chandola, T., & Marmot, M. (2003). Socio-economic status and health: causality and pathways. *Journal of Econometrics, 112* (1), 57-63.

Alam, K., & Mahal, A. (2014). Economic impacts of health shocks on households in low and middle income countries: a review of the literature. *Globalization and Health, 10* (21).

Alderman, H., & Paxson, C. H. (1992). Do the poor insure?: a synthesis of the literature on risk and consumption in developing countries. *World Bank Publications, 164*.

Alderman, H., Behrman, J. R., Kohler, H.-P., Maluccio, J. A., & Watkins, S. (2001). Attrition in longitudinal household survey data: some tests for three developing-country samples. *Demographic Research, 5*, 79-124.

Ambrosius, C., & Cuecuecha, A. (2013). Are remittances a substitute for credit? Carrying the financial burden of health shocks in national and transnational households. *World Development, 46*, 143-152.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Asfaw, A., & Von Braun, J. (2004). 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.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46* (3), 399-424.

Babiarz, P., Widdows, R., & Yilmazer, T. (2013). Borrowing to cope with adverse health events: liquidity constraints, insurance coverage, and unsecured debt. *Health Economics, 22* (10), 1177-1198.

Bai, C.-E., & Wu, B. (2014). Health insurance and consumption: evidence from China's new cooperative medical scheme. *Journal of Comparative Economics, 42* (2), 450-469.

Baker, M., Stabile, M., & Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources, 39* (4), 1067-1093.

Bender, K. A., & Theodossiou, I. (2009). Controlling for endogeneity in the health-socioeconomic status relationship of the near retired. *The Journal of Socio-Economics, 38* (6), 977-987.

Berman, R., Quinn, C., & Paavola, J. (2013). *Identifying drivers of household coping strategies to multiple climatic hazards in Western Uganda: implications for adapting to future climate change.* University of Leeds, Sustainability Research Institute. Centre for Climate Change Economics and Policy Working Paper n. 149.

Bloom, D. E., Cafiero, E. T., McGovern, M. E., Prettner, K., Stanciole, A., Weiss, J., et al. (2013). The economic impact of non-communicable disease in China and India: estimates, projections, and comparisons. *National Bureau of Economic Research, w19335*.

Blumenthal, D., & Hsiao, W. (2005). Privatization and its discontents — The Evolving Chinese Health Care System. *The New England Journal of Medicine, 353*, 1165-1170.

Bonfrer, I., & Gustafsson-Wright, E. (2016). Health shocks, coping strategies and foregone healthcare among agricultural households in Kenya. *Global Public Health*, *19*, 1-22.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources, 26* (1), 106-138.

Bryant, J. (1969). *Health and the Developing World.* Ithaca: Cornell University Press.

Burström, B., & Fredlund, P. (2001). Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Healt*h, *55* (11), 836-840.

Butler, J. S., Burkhauser, R. V., Mitchell, J. M., & Pincus, T. P. (1987). Measurement error in self-reported health variables. *The Review of Economics and Statistics*, *64* (9), 644-650.

Cai, L. (2010). The relationship between health and labour force participation: Evidence from a panel data simultaneous equation model. *Labour Economics, 17* (1), 77-90.

Cai, L., & Kalb, G. (2006). Health status and labour force participation: Evidence from Australia. *Health Economics, 15* (3), 241-261.

Case, A., & Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In D. A. Wise, *Analyses in the Economics of Aging* (pp. 185-211). Chicago: University of Chicago Press.

Chamon, M., & Prasad, E. (2008). Why are Saving Rates of Urban Households in China Rising? *NBER Working Paper, 14546*.

Chen, C., Ding, S., Cook, S., & Pong, M. (2014). *Coming Home: The Return of Migrant Workers with Illness or Work-Related Injuries in China's Hubei and Sichuan Provinces.* Geneva: The United Nations Research Institute for Social Development.

Chen, G., Liu, G. G., & Xu, F. (2014). The impact of the Urban Resident Basic Medical Insurance on health services utilisation in China. *Pharmacoeconomics, 32* (2), 277-292.

Chetty, R., & Looney, A. (2006). Consumption smoothing and the welfare conse- quences of social insurance in developing countries. *Journal of Public Economics, 90* (12), 2351–2356.

China Centers for Disease Prevention and Control. (2012). *China Health Statistics Summary.* China Centers for Disease Prevention and Control.

Clarke, P. M., Hayes, A. J., Glasziou, P. G., Scott, R., Simes, J., & Keech, A. C. (2009). Using the EQ-5D index score as a predictor of outcomes in patients with type 2 diabetes. *Medical Care*, *47*(1), 61-68.

Coile, C. C. (2004). Health shocks and couples' labor supply decisions. *National Bureau of Economic Research*. Working paper w10810.

Coile, C., & Milligan, K. (2009). How household portfolios evolve after retirement: The effect of aging and health shocks. *Review of Income and Wealth, 55* (2), 226-248.

Conning, J., & Udry, C. (2007). Rural financial markets in developing countries. *Handbook of Agricultural Economics, 3*, 2857-2908.

Cooper, C. L., & Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology, 49* (1), 11-28.

Currie, J., & Madrian, B. C. (1999). Health, health insurance and the labor market. In Ashenfelter, O. & Card, D., *Handbook of Labor Economics*, *3 (pp. 3309–416)*. Amsterdam: Elsevier Science.

Dano, A. M. (2005). Road injuries and long-run effects on income and employment. *Health Economics, 14* (9), 955-970.

Deaton, A. (1997). *The analysis of household surveys: A microeconometric approach to development policy.* Baltimore: John Hopkins University Press.

Démurger, S., & Xu, H. (2011). Return migrants: The rise of new entrepreneurs in rural china. *World Development, 39* (10), 1847–1861.

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *The World Bank Research Observer, 17* (2), 141-166.

Dercon, S., & Krishnan, P. (2000). In sickness and in health: Risk sharing within households in rural Ethiopia. *Journal of Political Economy, 108* (4), 688-727.

Dhanaraj, S. (2016). Economic vulnerability to health shocks and coping strategies: evidence from Andhra Pradesh, India.". *Health Policy and Planning*, *31*(6):749-58

Dodd, R., & Munck, L. (2001). *Dying for change: poor people's experience of health and ill-health.* Geneva: World Health Organization.

Dreger, C., Wang, T., & Zhang, Y. (2015). Understanding Chinese consumption: The impact of Hukou. *Development and Change, 46* (6), 1331-1344.

Eggleston, K., Ling, L., Qingyue, M., Lindelow, M., & Wagstaff, A. (2008). Health service delivery in China: a literature review. *Health Economics*, *17*(2), 149-165.

Fletcher, J. M., & Sindelar, J. L. (2009). *Estimating causal effects of early occupational choice on later health: Evidence using the psid.* National Bureau of Economic Research. Working Paper w15256

Flores, G., Krishnakumar, J., O'Donnell, O., & Van Doorslaer, E. (2008). Coping with health-care costs: implications for the measurement of catastrophic expenditures and poverty. *Health economics*, *17*(12), 1393-1412.

García-Gómez, P. (2011). Institutions, health shocks and labour market outcomes across Europe. *Journal of Health Economics, 30* (1), 200-213.

García-Gómez, P., Van Kippersluis, H., O'Donnell, O., & Van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, *48* (4), 873-909.

Genoni, M. E. (2012). Health shocks and consumption smoothing: Evidence from Indonesia. *Economic Development and Cultural Change, 60* (3), 475-506.

Gertler, P., & Gruber, J. (2002). Insuring consumption against illness. *American Economic Review, 92* (1), 51-70.

Gertler, P., Levine, D. I., & Moretti, E. (2009). Do microfinance programs help families insure consumption against illness? *Health Economics, 18* (3), 257-273.

Grimm, M. (2010). Mortality shocks and survivors' consumption growth. *Oxford Bulletin of Economics and Statistics, 72*, 146–171.

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy, 80* (2), 223–255.

Gupta, N. D., Kleinjans, K. J., & Larsen, M. (2015). The effect of a severe health shock on work behavior: Evidence from different health care regimes. *Social Science & Medicine, 136*, 44-51.

Gustafsson-Wright, E., Janssens, W., & van der Gaag, J. (2010). The inequitable impact of health shocks on the uninsured in Namibia. *Health Policy and Planning, 26(2):142-56*.

Hafez, R. (2014). *The impact of health insurance on financial risk protection in Ningxia, China .* University of Oxford, Nuffield Department of Population Health, Green Templeton College. Oxford: University of Oxford.

Hagan, R., Jones, A. M., & Rice, N. (2008). Health shocks and the hazard rate of early retirement in the ECHP. *Swiss Journal of Economics and Statistics, 144* (3), 323-335.

Halla, M., & Zweimüller, M. (2013). The effect of health on earnings: Quasi-experimental evidence from commuting accidents. *Labour Economics, 24*, 23-38.

Heltberg, R., & Lund, N. (2009). Shocks, coping, and outcomes for Pakistan's poor: health risks predominate,. *The Journal of Development Studies, 45* (6), 889-910.

Heltberg, R., Hossain, N., Reva, A., & Turk, C. (2012). Anatomy of coping: Evidence from people living through the crises of 2008-11. *World Bank Policy Research Working Paper, 5957*.

Islam, A., & 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.

Jialu, L. (2011). Human capital, migration and rural entrepreneurship in China. *Indian Growth and Development Review, 4* (2), 100-122.

Jiménez-Martín, S., Labeaga, J. M., & Vilaplana Prieto, C. (2006). A sequential model of older workers' labor force transitions after a health shock. *Health Economics, 15* (9), 1033-1054.

Jo, C. (2014). Cost-of-illness studies: concepts, scopes, and methods. *Clinical and Molecular Hepatology, 20* (4), 327-337.

Johnson, E., & Reynolds, C. L. (2013). The effect of household hospitalizations on the educational attainment of youth. *Economics of Education Review, 37*, 165-182.

Jones, A. M., Rice, N., & Zantomio, F. (2015). *Acute health shocks and labour market exits.* N/A.

Juwei, Z. (2010). Social protection in China: current status and challenges. In M. G. Asher, S. Oum, & F. Parulian, *Social Protection in East Asia - Current State and Challenges* (pp. 371-398). Jakarta: Economic Research Institute for ASEAN and East Asia.

Kapteyn, A., Smith, J. P., & Van Soest, A. (2011). Work disability, work, and justification bias in Europe and the United States. In D. A. Wise,

Explorations in the Economics of Aging (pp. 269-312). Chicago: University of Chicago Press.

Karlan, D., & Zinman, J. (2009). Expanding credit access: Using randomized supply decisions to estimate the impacts. *Review of Financial Studies , 2* (1), 433–64.

Kim, H., Yoon, W., & Zurlo, K. A. (2012). Health shocks, out-of-pocket medical expenses and consumer debt among middle-aged and older americans. *Journal of Consumer Affairs, 46* (3), 357-380.

Kuczmarski, M. F., Kuczmarski, R. J., & Najjar, M. (2001). Effects of age on validity of self-reported height, weight, and body mass index: Findings from the third national health and nutrition examination survey, 1988–1994. *Journal of the American Dietetic Association, 101* (1), 28-34.

Li, J., Zheng, G., Hui, L., Wang, L., Tang, Z., Shi, H., et al. (N.A.). *Situation Analysis of Ningxia Province (report by the China Climate Change Partnership).* Food and Agriculture Organization. Food and Agriculture Organization.

Li, X., & Freeman, R. B. (2015). How does china's new labour contract law affect floating workers? *British Journal of Industrial Relations, 53* (4), 711-735.

Li, Y., Wu, Q., Liu, C., Kang, Z., Xie, X., Yin, H., et al. (2014). Catastrophic health expenditure and rural household impoverishment in China: what role does the new cooperative health insurance scheme play? *Plos One, 9* (4).

Lindelow, M., & Wagstaff, A. (2005). *Health shocks in China: are the poor and uninsured less protected?* World Bank. Washington, D.C.: World Bank Policy Research.

Liu, J. X. (2010). *Healthy time, home production, and labor supply: the effects of health shocks on time use in China.* University of California, Health Services and Policy Analysis. Berkeley: University of California.

Liu, K. (2016). Insuring against health shocks: Health insurance and household choices. *Journal of Health Economics, 46*, 16–32.

Lunt, M. (2014). Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *American Journal of Epidemiology, 179* (2), 226-235.

Lutz, W., Sanderson, W., & Scherbov, S. (2008). The coming acceleration of global population ageing. *Nature*, 451 (7179), 716-719.

Mace, B. J. (1991). Full insurance in the presence of aggregate uncertainty. *Journal of Political Economy, 99* (5), 928-956.

Margolis, R. (2013). Health shocks in the family gender differences in smoking changes. *Journal of Aging and Health (2013) , 25* (5), 882-903.

McClellan, M. B. (1998). Health events, health insurance, and labor supply: Evidence from the health and retirement survey. In D. Wise, *Frontiers in the Economics of Aging* (pp. 301-350). Chicago: University of Chicago Press:.

McGeary, K. A. (2009). How do health shocks influence retirement decisions? *Review of Economics of the Household, 7* (3), 307-321.

McIntyre, D., Thiede, M., Dahlgren, G., & Whitehead, M. (2006). What are the economic consequences for households of illness and of paying for health care in low-and middle-income country contexts ? *Social Science & Medicine, 62* (4), 858-865.

Miguel, E. (2005). Poverty and witch killing. *The Review of Economic Studies, 62* (4), 1153-1172.

Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., & Urponen, H. (1997). Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology, 50* (5), 517-528.

Millán-Calenti, J. C., Tubío, J., Pita-Fernández, S., González-Abraldes, I., Lorenzo, T., Fernández-Arruty, T., et al. (2010). Prevalence of functional disability in activities of daily living (ADL), instrumental activities of daily living (IADL) and associated factors, as predictors of morbidity and mortality. *Archives of Gerontology and Geriatrics, 50* (3), 306-310.

Mitra, S., Palmer, M., Mont, D., & Groce, N. (2016). Can households cope with health shocks in vietnam? *Health Economics, 25(7):888-907.*

Mohanan, M. (2013). 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.

Morduch, J. (1995). Income smoothing and consumption smoothing. *The Journal of Economic Perspectives, 9* (3), 103-114.

Narayan, D., Patel, R., Schafft, K., Rademacher, A., & Koch-Schulte, S. (2000). *Voices of the poor: can anyone hear us?* New York: Oxford University Press.

National Bureau of Statistics. (2013). *China Statistical Yearbook.* Beijing: China Statistics Press.

National Bureau of Statistics of China. (2011, 04 29). *Communiqué of the National Bureau of Statistics of People's Republic of China on Major Figures of the 2010 Population Census.* Retrieved 03 25, 2016 from National Bureau of Statistics of China: http://www.stats.gov.cn/english/NewsEvents/201104/t20110429_26450.html

National Bureau of Statistics China. (2015, 02 26). *Statistical Communiqué of the People's Republic of China on the 2014 National Economic and Social Development*. Retrieved 02 18, 2016 from National Bureau of Statistics China: http://www.stats.gov.cn/english/PressRelease/201502/t20150228_687439.html

Ningxia Provincial Bureau of Statistics. (2013). *Nigxia Statistical Yearbook.* China Statistics Press.

O' Donnell, O., Van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). *analyzing health equity using household survey data: A guide to techniques and their implementation.* Washington, DC: The World Bank.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation, 9* (6), 1-12.

Ping, H., & Zhang, S. (2008). Migrant workers' remittances and rural development in China. N/A

Pohl, V., Neilson, C., & Parro, F. (2014). Health shocks, education, and labor market outcomes. N/A

Popkin, B. M., Du, S., Zhai, F., & Bing, Z. (2009). Cohort Profile: The China Health and Nutrition Survey—monitoring and understanding socio-economic and health change in China, 1989–2011. *International Journal of Epidemiology, 39*(6) 1435-1440.

Powell-Jackson, T., & Hoque, M. E. (2012). Economic consequences of maternal illness in rural Bangladesh. *Health Economics, 21* (7), 796-810.

Qi, Y., & Niu, J. (2013). Health selection effects in China's internal migration. *Asian Population Studies*, *9* (2), 142-155.

Qiu, P., Yang, Y., Zhang, J., & Ma, X. (2011). Rural-to-urban migration and its implication for new cooperative medical scheme coverage and utilization in China. *BMC Public Health*, *11* (1), 1.

Ramos-Goni, J. M., & Rivero-Arias, O. (2011). eq5d: A command to calculate index values for the EQ-5D quality-of-life instrument. *The Stata Journal, 11* (1), 120–125.

Riphahn, R. T. (1999). Income and employment effects of health shocks A test case for the German welfare state. *Journal of Population Economics, 12* (3), 363-389.

Sauerborn, R., Adams, A., & Hien, M. (1996). Household strategies to cope with the economic costs of illness. *Social Science & Medicine, 43* (3), 291-301.

Sparks, K., Cooper, C., Fried, Y., & Shirom, A. (1997). The effects of hours of work on health: a meta-analytic review. *Journal of Occupational and Organizational psychology, 70* (4), 391-408.

Sparrow, R., Van de Poel, E., Hadiwidjaja, G., Yumna, A., Warda, N., & Suryahadi, A. (2014). Coping with the economic consequences of ill health in Indonesia. *Health Economics, 23* (6), 719-728.

Stern, S. (1989). Measuring the effect of disability on labour force participation. *The Journal of Human Resources, 24* (3), 361-395.

Sun, X., Jackson, S., Carmichael, G., & Sleigh, A. C. (2009). Catastrophic medical payment and financial protection in rural China: evidence from the New Cooperative Medical Scheme in Shandong Province. *Health Economics, 18* (1), 103-19.

Sundmacher, L. (2012). The effect of health shocks on smoking and obesity. *The European Journal of Health Economics, 13* (4), 451-460.

Tong, Y., & Piotrowski, M. (2012). Migration and health selectivity in the context of internal migration in China, 1997–2009. *Population Research and Policy Review, 31* (4), 497-543.

Uchimura, H., & Jütting, J. P. (2009). Fiscal decentralization, Chinese style: good for health outcomes? *World Development, 37* (12), 1926-1934.

Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria. *The Review of Economic Studies, 61* (3), 495-526.

Valle, G., Weeks, J. A., Taylor, M. G., & Eberstein, I. W. (2013). Mental and physical health consequences of spousal health shocks among older adults. *Journal of Aging and Health, 25* (7), 1121-1142.

Vermeulen, F. (2002). Collective household models: principles and main results. *Journal of Economic Surveys, 16* (4), 533-564.

Wagstaff, A. (2002). Poverty and health sector inequalities. *Bulletin of the World Health Organization, 80* (2), 97-105.

Wagstaff, A. (2007). The economic consequences of health shocks: evidence from Vietnam. *Journal of Health Economics, 26* (1), 82–100.

Wagstaff, A., & Lindelow, M. (2014). Are health shocks different? evidence from a multishock survey in Laos. *Health Economics, 23* (6), 706-718.

World Bank. (2009). *China: From Poor Areas to Poor People: China's Evolving Poverty Reduction Agenda—An Assessment of Poverty and Inequality. Report 47349-CN.* Poverty Reduction and Economic Management Department, East Asia and Pacific Regions. Washington: World Bank.

World Bank. (2016). *World DataBank*. Retrieved 03 25, 2016 from World Bank: http://databank.worldbank.org/data/

Xie, Y., & Hu, J. (2014). An introduction to the China family panel studies (CFPS). *Chinese Sociological Review, 47* (1), 3-29.

Yi, J., Heckman, J. J., Zhang, J., & Conti, G. (2015). Early Health Shocks, Intra-household Resource Allocation and Child Outcomes. *The Economic Journal, 125* (588), F347-F371.

Yilma, Z., Mebratie, A., Sparrow, R., Abebaw, D., Dekker, M., Alemu, G., et al. (2014). Coping with shocks in rural Ethiopia. *Journal of Development Studies, 50* (7), 1009-1024.

Yip, W. C., Hsiao, W. C., Chen, W., Hu, S., Ma, J., & Maynard, A. (2012). Early appraisal of China's huge and complex health-care reforms. *The Lancet, 379* (9818), 833-842.

Zhang, L., Liu, S., Zhang, G., & Wu, S. (2015). Internal migration and the health of the returned population: a nationally representative study of China. *BMC public Health, 15* (1).

Zhang, X., Zhou, X., Huang, X., Miao, S., Shan, H., Jing, S., et al. (2014). The Analysis of the Disease Spectrum in China. *BioMed Research International*, *16* (12): 28812–28840.

Zucchelli, E., Jones, A. M., Rice, N., & Harris, A. (2010). The effects of health shocks on labour market exits: Evidence from the HILDA survey. *Australian Journal of Labour Economics, 13* (2), 191-218.

APPENDIX A – QUESTIONNAIRES
(Sample of questions used in the analyses only, not full questionnaires)

Appendix A1 – Questions used from the Ningxia Project Household Survey questionnaire

Variable	Question as it appears in the questionnaire	Variable name	Notes
Household head	Relationship to head of household: (1) Head of household (2) Spouse (3) Son or daughter (4) Grandchild (5) Parent (6) Grandparent (7) Sibling (8) Other	A2	
Gender	Gender: (1) Male (2) Female	A3	
Ethnicity	Ethnicity: (1) Han (2) Meng (3) Hui (4) Zang (5) Wei (6) Miao (7) Other	A4	
Age	Age (based on full years of life)	A5	
Marital status	Marital status (1) unmarried (2) married (3) divorced (4) widow/widower (5) other	A6	
Educational attainment	Level of education: (1) Never attended school (2) Elementary school (3) Middle school (4) High school or more	A7	
Main occupation	Main occupation: (1) Farmer (2) Labourer (unskilled) (3) Village cadre (4) Village doctor (5) Tradesperson (6) Teacher (7) Student (8) Business owner (9) Unemployed (10) Other **(list)**	A8	
Migration status	In the past year since this survey, have you left town to work as a migrant? (1) Yes (answer A9.1.1-A9.1.9) (2) No (Skip to A9.2)	A9	
Migration	In the past year, where did you go as a migrant ?	A9.1.1	Only available in 2015

location	(1) In rural area of another county but same province (2) In rural area of another province: Name of location/province (3) In urban area of another county but same province (4) In urban area of another province: Name of location/province		
Industry migration	Which industry did you work? (1) Service industry (2)Manufacture (3)Agriculture (4)Construction work (5) Others	A9.1.2	Only available in 2015
First migration year	When did you first migrate for the first time (Year)?	A1.3.1	Only available in 2015
Previous migration location	Please state where you went in previous years (from 2008 to 2013)	A1.3.2 (A9.1.3.2.1 - A9.1.3.2.6)	Only available in 2015
Migration alone	Did you migrate on your own? (1) Yes, on my own (2) No, I migrated with other fellow villagers (3) I migrated with other family members	A9.1.4	Only available in 2015
Migration months	In the past year approximately how many months total did you spend working out of town as a migrant? (unit: months)	A9.1	Available all years
Return reason	If migrated less than six months in the preceding year, why did you return? (1) Harvest (2) Family problems (3) Health problems (4) Have another job (not agricultural) in hometown (5) Unemployed in the city	A9.1.5	Only available in 2015
Returned migrant	Have you ever lived away from your home to work as a migrant worker in your life? (1) Yes (2) No	A9.2	Only available in 2015
Migration return reason	If yes, why are you not migrating anymore? (1) Looking after other family members (including elderly and child) (2) I am too old (3) My health is not good (4) Job prospects were poor	A9.2.1	Only available in 2015

	(5) I didn't like working in the city (6) Others: please state reason		
Hospital admission times	In the past year, how many times were you hospitalized?	C3	
Diagnosis admission 1	What is the name of the illness (injury, poison, etc.) for which you were hospitalized? **(During the survey write the name of the ailment, during the cross-check fill in the code for the ailment)**	C4	
Referral	Were you referred to the hospital this time? (1) Yes (2) No	C4.3	
Diagnosis admission 2	What is the name of the illness (injury, poison, etc.) for which you were hospitalized? **(During the survey write the name of the ailment, during the cross-check fill in the code for the ailment)**	C5	
Chronic disease	Have you suffered from a chronic disease that was **formally diagnosed** by a doctor? (1) Yes (2) No	D1	
Eq5d : item 1	Today, in terms of your mobility: (1) Can move in four directions without any difficulty (2) Movement is a little difficult (3) Bedridden	E1	
Eq5d : item 2	Today, in terms of your ability to take care of yourself (wash hands and face, get dressed): (1) No problems at all (2) Some problems (3) Unable to wash or dress oneself	E2	
Eq5d : item 3	Today, in terms of your ability to carry out regular daily activities (work, read, or household chores) : (1) Can carry out daily activities without any problem (2) Some	E3	

	problems (3) Unable to carry out daily activities		
Eq5d : item 4	Today, in terms of pain or discomfort: (1) No pain or discomfort (2) Moderate pain or discomfort (3) Extreme pain or discomfort	E4	
Eq5d : item 5	Today, in terms of your level of worry or depression : (1) No worry or depression (2) Moderate worry or depression (3) Extreme worry or depression	E5	
Self-rated health	Generally speaking, your health is : (1) Extremely good (2) Good (3) Fine (4) Average (5) Poor		
Assets	Does your family own any of the following possessions? • Watch/Alarm clock • Bicycle • Radio • Black & White TV • Color TV • Sewing machine • Motorbike • Car • Electric refrigerator • Washing Machine • Telephone (including cell phone) • Farming Machine • DVD/VCD/soundsystem • Camera/Video camera • Air conditioning	J1.1 – J1.21	Respondent tick yes or no for each item, with the exception of cow, sheep, horse, donkey, mule, pigs for which the number of animals is provided.

	• Cow • Sheep • Horse/Donkey/Mule • Pigs • Irrigated Land • Land on a Mountain		
House Quality: q1	What type of house does your family live in? (1) Brick, earth, and mud (2) Wooden blocks (3) Earth and wood (4) All brick (5) Cave dwelling (6) Other		
House Quality: q2	What material is the floor of your house made of? (1) Dirt (2) Brick (3) Ceramic tile (4) Wooden planks (5) Laminate (6) Other	J2	
Water	The primary source of your family's drinking water is: (1) Tap water (2) Spring water (3) Hand-drawn well water (4) Cellar water (5) Well water (6) River or lake water (7) Ditch water (8) Other	J3	
Toilet	What type of toilet does your family have (1) Flush toilet (2-6) have toilet but not flush toilet (7) No toilet (8) Other *[note: (2)(3)(4)(5)(6) were not translated as they do not have an English equivalent]*	J4	
Fuel cooking	What type of fuel does your family typically use for cooking? **(can select up to two)** : (1) Coal (2) Electric (3) Kerosene (4) Natural gas (5) Wood/hay/etc. (6) Charcoal (7) Other	J5	
Loan dummy	Does your family have any outstanding loans? (1)Yes (2) No (**skip to J11**) (3) Not sure	J10	

Loans reason	If your family has loans, what was the main reason for taking out the loan? (1) Purchase food (2) Build or fix a house (3) See the doctor (4) Send the children to school (5) Pay taxes (6) Wedding or funeral expenses (7) Agricultural production (8) Other **(list)**	J10.1	
Loan amount	What is the total amount of the loan?	J10.1.1	
Consumption: monthly items	***Average monthly expenditure on the following products*** • Grains, meat, fruit, vegetables, etc. • Oil, salt, soy sauce, vinegar, tea, seasoning, etc. • Daily commodities, such as soap, paper, pens, newspaper, etc. • Utilities, such as electric, water, heating, cooking fuel, etc. • Rent • Cigarettes, alcohol, etc. • Transportation, phone, postage, etc. • Cultural and entertainment activities	J13.1 - J13.8	
Consumption: yearly items	***Expenditure on the following commodities over the past year (12 months)*** • Clothing • Education: tuition, books, school supplies (**including living expenses for students not living at home**) • New Cooperative Medical Scheme premium • Medicine, medical expenses, etc. **(only includes what you personally paid for health services and medicine)** • Gifts, such as a dowry, wedding gifts, New Year's gifts, funeral expenses, gifts to friends and family • Entertainment, gambling, etc.	J14.1-J12.11	

	• Taxes, etc. • Insurance expenditure **(not including** integrated health insurance for urban and rural areas**)** • Non-durable goods, such as TV, radio, fan, bicycle, motorbike, car, etc. • Building and/or repairing buildings • Buying apartment or house • Other		
Savings	Over the past year was your family's income sufficient? How much were you able to save? (RMB) **(If none write "0".)**	J15	
Home produced goods	The value of total household consumption of home-produced items over the past year • Wheat • Corn • Vegetables • Meat • Eggs • Fruits • Buckwheat • Flax • Potatoes • Other	J16.1 – J16.10	For each item, respondent report: Total quantity produced, Quantity sold, Quantity consumed, Market price & Total value
Income	In the past year, total household income (RMB) **(including money sent home by family members working out of town as migrants)**	J17	
Remittances	Total amount of remittances sent home by family members working out of town as migrants (RMB)	J17.1	

Appendix A2 – Questions used from the CFPS (household survey questionnaire)

NB: the variable names are given for 2010 only, please bear in mind that the consistency between the two data waves is not as good as for the Ningxia Project data, especially with regards to work time.

Variable	Question as it appears in the questionnaire	Variable name (from 2010 questionnaire)	Notes
Urban	Type of village/residential community 1. City 2. Town 3. Rural village	Z7	This variable was recoded as urban for response 1 and 2 and rural for 3.
Household size	How many people besides yourself are currently living together in this family?	Part I - A3	Household size was calculated from this question by adding one to the reported figure
Age	Age _____	Part I - B1	
Gender	Family member' sex	Part I - B2	
Marital status	Family member's marital status: 1. Never married 2. Married 3.Cohabitation 4. Divorced 5. Widowed	Part I - B3	
Educational attainment	Family member's highest level of education completed 1. Illiterate/Semi-literate 2. Primary school 3. Middle school 4. High school 5. 2- or 3-year college 6. 4-year college/Bachelor's degree 7. Master's degree 8. Doctoral degree	Part I - B4	This question was re-coded with the following categories: No school, primary school, middle school, high school or more
Household head	Who is the head of the household? [single response]	Part I - B7	

Water	What kind of water does your family normally use for cooking? 1. River/Lake water 2. Well/Spring water 3. Tap water 4. Mineral/Purified/Filtered water 5. Rainwater 6. Cellar water 7. Pond water 77. Other _____ [Please specify]	Part II – B1	This variable was recoded as 'tap and filtered water' (1) and other water (0)
Electricity	What is the electricity condition at your home? 1. No electricity 2. Frequent power outage 3. Occasional power outage 4. No power outage at all	Part II – B3	This variable was recoded as poor electricity, medium electricity and good electricity
Toilet	What kind of restroom/toilet facilities does your family have? 1. Indoor flush toilet 2.Outdoor private flush toilet 3. Outdoor public flush toilet 4. Indoor non-flush toilet 5.Outdoor private non-flush toilet 6. Outdoor public non-flush toilet	Part II – B5	
Consumption: monthly (last month used as a recall period)	Last month the expenditure on: • Food • Daily used commodities and necessities • Transportation (including expenditures for car maintenance) • Communication • Family member support • Home mortgage • Car mortgage • Other mortgages • House rental	Part II- H301 - 309	Respondent choose from one of those figures. Consumption at the level of the community was compiled rather than using those figures directly
Consumption: yearly (last year used as a recall	Last year, the expenditure on: • Electricity • Medical care	Part II – H401 - 411	

period)	• Clothing • Education • Cultural, entertainment and leisure activities • Living expense (community management, heating cost etc.) • Miscellaneous goods and services • Commercial insurance • Family's own marriages and funerals • Other items		
Agricultural work	Are you currently primarily engaging in agricultural work?	G4	
Type of crops	Last year, did your family grow any of the following agricultural product? 1. Rice 2. Wheat 3. Corn 4. Cotton 5. Forestry products/Forestry side line products 78. None of these	K6	This question does not appear in 2012
Hospital admission	How many times were you hospitalized due to illness last year? _____times	P601	
Hospital length	P603 How many days did you stay in the hospital?	P603	
Wage work: months	How many months did you work last year? _____months	G401	
Wage work: days	How many days per month on average did you work during the working months in the last year? _____days	G402	
Wage work: hours	How many hours per day did you work during the working days in the last year?_____	G403	
Agricultural work: months	How many months did you engage in agricultural work last year? ____ months	H1	In 2012, this question was rephrased and split in two between the busy season and the off farm season or 'slack farming season'

			(G202-G20)
Agricultural work: days	How many days per month on average did you work during the months when you engaged in agricultural work last year? _____ days	H2	In 2012, this question was rephrased and split in two between the busy season and the off farm season (G203)
Agricultural work: hours	How many hours per day on average did you work during the months when you engaged in agricultural work last year? _____ hours	H3	In 2012, this question was rephrased and split in two between the busy season and the off farm season (G203-G205)
Complementary non-agricultural work (waged)	Are you currently engaging in any paid non-agricultural work?	H4	In 2012, this question was rephrased and split in two between the busy season and the off farm season (G204-G206)
Complementary non-agricultural work (waged)	When did you start this non-agricultural work? ____ year ____ month	H409	This was used and combined with the interview date to calculate the number of months in the last 12 months the respondent was engaged in this activity
Complementary non-agricultural work (waged)	How many hours per day on average do you spend per day on this non-agricultural work? _____ hours	H410	The number of hours was multiplied using a 7 day week

Appendix B. Full regression tables for Chapters 4-7

Appendix B1 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on household income

	Income						
Hospital shock	**271.5**						
	(328.4)						
Hhead shock		**-999.8****					
		(408.5)					
Shock man			**-948.8***				
(working age)			**(515.6)**				
Shock woman				**-193.0**			
(working age)				**(410.2)**			
Shock elderly					**188.7**		
					(396.1)		
Severe shock						**441.4**	
						(383.0)	
Multi shocks							**-404.1**
							(933.1)
House index	5,291***	5,290***	5,290***	5,293***	5,292***	5,292***	5,292***
	(204.5)	(204.4)	(204.5)	(204.6)	(204.5)	(204.4)	(204.5)
Hhead educ	3,281***	3,271***	3,268***	3,280***	3,280***	3,279***	3,279***
	(450.8)	(450.7)	(450.7)	(450.4)	(450.5)	(451.0)	(450.5)
Hhead marital	-694.3	-717.9	-690.9	-695.0	-693.9	-698.8	-696.3
	(845.6)	(845.6)	(844.7)	(845.5)	(845.5)	(846.0)	(845.4)
Hhead labourer	1,156	1,186	1,173	1,170	1,161	1,158	1,171
	(727.6)	(727.0)	(726.6)	(726.8)	(727.8)	(726.4)	(726.7)
Hhead farmer	-2,128***	-2,110***	-2,119***	-2,123***	-2,128***	-2,125***	-2,120***
	(697.0)	(696.7)	(696.5)	(696.5)	(697.5)	(696.7)	(696.8)
Hhead unemployed	-4,284***	-4,202***	-4,257***	-4,263***	-4,279***	-4,288***	-4,253***
	(1,030)	(1,032)	(1,031)	(1,028)	(1,031)	(1,027)	(1,028)
Hhead gender	2,672*	2,652*	2,668*	2,674*	2,680*	2,666*	2,671*
	(1,462)	(1,461)	(1,459)	(1,460)	(1,460)	(1,462)	(1,460)
Hhead age	622.6***	627.0***	623.4***	623.2***	622.9***	622.2***	623.4***
	(48.42)	(48.22)	(48.29)	(48.29)	(48.38)	(48.33)	(48.32)
Nb of migrants	783.6***	773.3***	776.4***	780.9***	783.5***	784.4***	780.9***
	(214.8)	(214.3)	(215.0)	(215.0)	(214.6)	(214.9)	(214.8)
HH size	1,481***	1,479***	1,493***	1,488***	1,484***	1,480***	1,487***
	(187.9)	(188.2)	(188.5)	(188.7)	(187.8)	(188.1)	(188.0)
Nb of children	-570.8**	-561.2**	-577.2**	-569.5**	-568.4**	-567.5**	-567.6**
	(236.7)	(237.2)	(236.7)	(237.2)	(236.7)	(236.8)	(236.6)
Nb of seniors	-123.2	-96.11	-113.7	-114.8	-126.7	-132.5	-106.0
	(347.8)	(349.2)	(348.7)	(348.4)	(347.0)	(348.6)	(349.0)
Observations	25,240	25,240	25,240	25,240	25,240	25,240	25,240
Number of HH	9,764	9,764	9,764	9,764	9,764	9,764	9,764

Hhead stands for 'household head' and HH for 'household'

Each column and line corresponds to a different regression. Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).

*** p<0.01, ** p<0.05, * p<0.1

Appendix B2 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on health expenditure

	Health expenditure						
Hospital shock	**1,827*** (108.9)**						
Hhead shock		**2,044*** (169.8)**					
Shock man (working age)			**2,245*** (211.1)**				
Shock woman (working age)				**1,607*** (143.0)**			
Shock elderly					**1,767*** (142.0)**		
Severe shock						**2,474*** (150.6)**	
Multi shocks							**2,624*** (414.3)**
House index	240.3***	252.7***	251.7***	244.9***	247.9***	242.3***	245.3***
	(50.82)	(51.37)	(51.50)	(51.10)	(51.86)	(50.30)	(51.57)
Hhead educ	189.6*	199.7*	216.9**	185.0*	180.3*	175.5*	190.9*
	(103.7)	(105.2)	(104.8)	(103.6)	(102.8)	(103.7)	(104.5)
Hhead marital	-41.89	1.472	-47.87	-38.10	-31.22	-56.77	-37.42
	(240.9)	(239.4)	(240.4)	(240.8)	(240.4)	(240.1)	(240.7)
Hhead labourer	160.0	175.1	209.5	202.6	184.7	185.7	197.6
	(151.5)	(152.3)	(150.2)	(152.6)	(152.8)	(152.4)	(153.0)
Hhead farmer	-34.82	-46.86	-20.14	-18.30	-32.88	-12.27	-35.17
	(141.1)	(141.3)	(140.2)	(143.0)	(143.2)	(143.8)	(143.7)
Hhead unemployed	223.4	217.0	314.6	325.5	251.8	216.6	274.6
	(280.1)	(278.4)	(278.4)	(281.8)	(280.0)	(281.1)	(282.4)
Hhead gender	297.7	366.4	356.8	357.8	385.1	276.5	363.1
	(310.7)	(301.8)	(313.0)	(302.4)	(306.5)	(306.9)	(311.1)
Hhead age	56.85***	54.27***	61.35***	62.14***	57.88***	55.59***	61.13***
	(8.478)	(8.729)	(8.746)	(8.743)	(8.546)	(8.615)	(8.796)
Nb of migrants	-86.78*	-74.63	-79.72*	-86.83*	-83.11*	-83.43*	-88.64*
	(46.51)	(47.14)	(46.67)	(46.80)	(47.47)	(46.84)	(46.82)
HH size	57.25	96.32**	67.31	72.64	78.97*	57.92	76.15
	(45.41)	(45.77)	(45.38)	(45.37)	(45.99)	(46.17)	(46.22)
Nb of children	-74.64	-70.01	-36.78	-53.38	-58.59	-55.17	-62.92
	(61.98)	(63.45)	(62.87)	(62.25)	(62.70)	(62.25)	(62.85)
Nb of seniors	555.4***	607.6***	641.5***	667.2***	502.7***	521.9***	603.2***
	(86.06)	(84.19)	(86.23)	(87.00)	(84.20)	(85.61)	(86.75)
Observations	26,375	26,375	26,375	26,375	26,375	26,375	26,375
Number of HH	9,891	9,891	9,891	9,891	9,891	9,891	9,891

Hhead stands for 'household head' and HH for 'household'

Each column and line corresponds to a different regression. Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).

*** p<0.01, ** p<0.05, * p<0.1

Appendix B3 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on savings and assets

	Savings	Savings	Savings	Assets	Assets	Assets
Hospital shock	**-144.8****			**-0.000**		
	(66.48)			**(0.00945)**		
Hhead shock		**-190.4*****			**-0.026***	
		(72.92)			**(0.013)**	
Multiple shocks			**-142.1**			**0.037**
			(179.1)			**(0.026)**
House index	307.8***	306.8***	307.4***	0.841***	0.841***	0.841***
	(42.35)	(42.28)	(42.29)	(0.006)	(0.006)	(0.006)
Hhead educ	274.0***	273.1***	273.4***	0.069***	0.069***	0.069***
	(85.34)	(85.32)	(85.33)	(0.013)	(0.013)	(0.013)
Hhead marital	177.8	172.8	177.2	-0.046*	-0.046*	-0.046*
	(146.8)	(147.1)	(146.6)	(0.025)	(0.025)	(0.025)
Hhead labourer	-11.88	-11.94	-14.92	-0.027	-0.027	-0.028
	(162.1)	(162.0)	(162.0)	(0.019)	(0.019)	(0.019)
Hhead farmer	-218.3	-216.3	-218.2	-0.041**	-0.040**	-0.041**
	(145.4)	(145.2)	(145.2)	(0.017)	(0.017)	(0.017)
Hhead unemployed	-147.3	-145.0	-152.4	-0.105***	-0.103***	-0.106***
	(169.1)	(168.1)	(168.6)	(0.029)	(0.029)	(0.029)
Hhead gender	157.4	151.1	154.0	0.043	0.043	0.044
	(332.9)	(334.1)	(333.0)	(0.045)	(0.045)	(0.045)
Hhead age	26.38***	26.65***	26.00***	0.013***	0.013***	0.013***
	(7.830)	(7.878)	(7.834)	(0.001)	(0.001)	(0.001)
Nb of migrants	113.1**	111.8**	113.5**	-0.019***	-0.02***	-0.019***
	(46.90)	(46.85)	(46.91)	(0.006)	(0.005)	(0.006)
HH size	-69.09**	-72.17**	-70.81**	0.052***	0.051***	0.051***
	(33.61)	(33.75)	(33.74)	(0.005)	(0.005)	(0.005)
Nb of children	126.3***	125.9***	124.9***	-0.057***	-0.057***	-0.057***
	(47.20)	(47.18)	(47.10)	(0.006)	(0.007)	(0.007)
Nb of seniors	-51.78	-55.79	-56.63	0.000	0.000	-0.000
	(66.45)	(65.83)	(65.94)	(0.012)	(0.012)	(0.012)
Observations	25,432	25,432	25,432	26,418	26,418	26,418
Number of HH	9,784	9,784	9,784	9,895	9,895	9,895

Hhead stands for 'household head' and HH for 'household'

Each column and line corresponds to a different regression. Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).

*** p<0.01, ** p<0.05, * p<0.1

Appendix B4 - Full regression tables for results presented in Chapter 4 on the impact of health shocks (by type of shock) on remittances and loans

	Remittances	Remittances	Remittances	Loans	Loans	Loans
Hospital shock	**567.6***			**2,100*****		
	(328.0)			**(476.2)**		
Hhead shock		**939.6***			**1,662****	
		(498.0)			**(644.5)**	
Multiple shocks			**1,025**			**4,476****
			(905.7)			**(1,788)**
House index	2,106***	2,115***	2,109***	3,578***	3,589***	3,580***
	(185.8)	(185.5)	(185.9)	-340.8	(340.3)	(341.1)
Hhead educ	961.6**	956.5**	959.2**	3,000***	3,005***	3,004***
	(416.0)	(415.3)	(414.4)	-623.4	(623.9)	(622.0)
Hhead marital	-460.4	-441.6	-456.1	-1064	-1,023	-1,056
	(825.6)	(821.2)	(823.0)	-1227	(1,229)	(1,221)
Hhead labourer	1,558*	1,601*	1,570*	1811	1,842	1,828
	(937.9)	(942.2)	(940.0)	-1206	(1,209)	(1,208)
Hhead farmer	-641.7	-602.5	-636.0	504.5	498.7	479.3
	(888.0)	(892.2)	(889.3)	-1129	(1,131)	(1,131)
Hhead unemployed	-110.8	-81.44	-87.67	-1185	-1,153	-1,178
	(1,367)	(1,366)	(1,369)	-1297	(1,297)	(1,299)
Hhead gender	2,673	2,607	2,699	1006	1,071	1,078
	(1,817)	(1,816)	(1,816)	-2238	(2,239)	(2,236)
Hhead age	245.5***	241.6***	246.7***	461.1***	460.9***	466.4***
	(45.27)	(45.18)	(45.31)	-63.66	(63.71)	(63.57)
Nb of migrants	1,951***	1,943***	1,959***	-1,353***	-1,347***	-1,352***
	(217.6)	(218.0)	(218.0)	-331.7	(331.8)	(330.4)
HH size	-78.95	-65.61	-72.79	1,749***	1,791***	1,761***
	(205.5)	(204.9)	(205.0)	-296.1	(296.5)	(298.0)
Nb of children	394.1	390.4	394.4	-1,412***	-1,405***	-1,400***
	(257.7)	(258.0)	(257.6)	-381.1	(381.5)	(381.6)
Nb of seniors	176.1	186.5	199.8	-740.9	-662.7	-705.0
	(414.5)	(412.6)	(414.8)	-600.4	(599.1)	(599.8)
Observations	11,819	11,819	11,819	26,539	26,539	26,539
Number of HH	6,313	6,313	6,313	9,915	9,915	9,915

Hhead stands for 'household head' and HH for 'household'

Each column and line corresponds to a different regression. Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).

*** p<0.01, ** p<0.05, * p<0.1

Appendix B5 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on total consumption

	(1)	(2)	(3)	(4)
		Total consumption		
Health shock	**2,551*** **			
	(516.6)			
Hhead shock		**2,542*** **		
		(694.6)		
Severe shock			**4,492*** **	
			(579.3)	
Multi shocks				**4,825*** **
				(1,591)
House index	8,437***	8,451***	8,438***	8,442***
	(390.7)	(390.2)	(389.7)	(391.3)
Hhead educ	3,285***	3,294***	3,263***	3,297***
	(611.1)	(613.3)	(610.8)	(613.0)
Hhead marital	-2,763**	-2,712**	-2,788**	-2,757**
	(1,219)	(1,220)	(1,214)	(1,218)
Hhead labourer	3,807***	3,839***	3,830***	3,836***
	(1,084)	(1,080)	(1,086)	(1,081)
Hhead farmer	1,913*	1,905*	1,948*	1,887*
	(1,023)	(1,020)	(1,025)	(1,021)
Hhead unemployed	-1,459	-1,438	-1,512	-1,437
	(1,421)	(1,413)	(1,424)	(1,415)
Hhead gender	2,951	3,038	2,875	3,084
	(2,179)	(2,180)	(2,169)	(2,179)
Hhead age	740.1***	738.3***	736.2***	747.0***
	(70.79)	(71.19)	(70.62)	(70.95)
Nb of migrants	-2,032***	-2,021***	-2,025***	-2,030***
	(302.0)	(300.8)	(300.6)	(301.0)
HH size	3,564***	3,619***	3,554***	3,585***
	(325.7)	(325.0)	(324.8)	(325.0)
Nb of children	-3,051***	-3,043***	-3,025***	-3,036***
	(443.0)	(443.4)	(443.0)	(442.8)
Number of senior in HH	-429.4	-340.3	-536.2	-369.2
	(560.9)	(560.5)	(560.2)	(562.7)
Observations	26,090	26,090	26,090	26,090
Number of HH	9,852	9,852	9,852	9,852

Hhead stands for 'household head' and HH for 'household'
Each column and line corresponds to a different regression (with household fixed effects). Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B6 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on NFNM consumption

	(1)	(2)	(3)	(4)
	NFNM consumption			
Health shock	**141.8**			
	(363.6)			
Hhead shock		**-348.7**		
		(495.7)		
Severe shock			**937.2****	
			(430.0)	
Multi shocks				**1,983**
				(1,228)
House index	5,748***	5,748***	5,747***	5,747***
	(282.2)	(282.1)	(282.0)	(282.1)
Hhead educ	2,243***	2,241***	2,240***	2,247***
	(434.2)	(434.2)	(433.7)	(434.1)
Hhead marital	-2,695***	-2,701***	-2,705***	-2,697***
	(863.6)	(863.9)	(863.0)	(864.0)
Hhead labourer	2,196**	2,207**	2,191**	2,180**
	(862.3)	(862.1)	(862.4)	(862.7)
Hhead farmer	741.8	748.3	746.7	723.0
	(808.6)	(808.5)	(808.5)	(809.3)
Hhead unemployed	-1,988*	-1,961*	-2,026*	-2,035*
	(1,061)	(1,059)	(1,063)	(1,062)
Hhead gender	1,630	1,629	1,607	1,650
	(1,460)	(1,459)	(1,459)	(1,458)
Hhead age	467.3***	468.9***	465.4***	467.6***
	(48.19)	(48.40)	(47.99)	(48.03)
Nb of migrants	-1,302***	-1,306***	-1,298***	-1,297***
	(225.5)	(225.6)	(225.6)	(225.6)
HH size	2,674***	2,675***	2,664***	2,666***
	(228.6)	(228.8)	(228.4)	(228.0)
Nb of children	-2,698***	-2,695***	-2,695***	-2,700***
	(299.2)	(299.5)	(299.3)	(298.9)
Number of senior in HH	-927.8**	-914.7**	-969.6**	-949.4**
	(375.4)	(374.5)	(374.3)	(377.6)
Observations	26,067	26,067	26,067	26,067
Number of HH	9,832	9,832	9,832	9,832

Hhead stands for 'household head' and HH for 'household'
Each column and line corresponds to a different regression (with household fixed effects). Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B7 - Full regression tables for results presented in Chapter 5 on the impact of health shocks (by type of shock) on food consumption

	(1)	(2)	(3)	(4)
		Food consumption		
Health shock	**69.66**			
	(105.4)			
Hhead shock		**-246.1****		
		(119.0)		
Severe shock			**96.78**	
			(113.5)	
Multi shocks				**269.6**
				(274.2)
House index	692.3***	691.4***	692.4***	692.3***
	(57.66)	(57.63)	(57.69)	(57.70)
Hhead educ	598.4***	597.5***	597.8***	599.3***
	(126.8)	(126.8)	(126.9)	(126.8)
Hhead marital	-255.2	-260.6	-256.3	-255.2
	(258.7)	(259.1)	(258.7)	(258.5)
Hhead labourer	523.1***	530.7***	524.1***	522.4***
	(195.2)	(194.9)	(195.2)	(195.4)
Hhead farmer	415.8**	420.9**	416.7**	414.0**
	(167.5)	(167.3)	(167.4)	(167.9)
Hhead unemployed	-19.81	-1.482	-20.86	-22.59
	(243.1)	(243.0)	(243.2)	(243.8)
Hhead gender	294.8	293.9	294.3	299.0
	(355.0)	(354.4)	(354.9)	(354.6)
Hhead age	99.43***	100.5***	99.37***	99.59***
	(11.43)	(11.48)	(11.43)	(11.41)
Nb of migrants	-211.1***	-214.0***	-211.1***	-210.8***
	(57.87)	(57.66)	(57.82)	(57.82)
HH size	521.5***	521.5***	521.6***	521.5***
	(54.68)	(54.87)	(54.68)	(54.90)
Nb of children	-97.76	-95.60	-97.14	-97.59
	(69.78)	(69.79)	(69.68)	(69.60)
Number of senior in HH	-55.19	-47.42	-56.82	-55.26
	(108.9)	(109.2)	(108.7)	(109.3)
Observations	26,065	26,065	26,065	26,065
Number of HH	9,868	9,868	9,868	9,868

Hhead stands for 'household head' and HH for 'household'
Each column and line corresponds to a different regression (with household fixed effects). Regressions on remittances are only performed on households with at least one migrant worker. Differences in sample size between regressions might arise from missing values or cleaning of the top 1% of outliers. All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B8 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on total consumption (using diff-in-diff strategy)

	(1)	(2)	(3)	(4)
		Total consumption		
Health shock 2009	**2,795*** **			
	(663.7)			
Health shock 2009-2011		**-4,479*** **		
		(1,377)		
Health shock 2009-2012			**-4,773*** **	
			(1,383)	
Health shock 2009-2015				**-7,672*** **
				(1,922)
House index	4,292***	4,910***	6,379***	12,278***
	(308.9)	(489.4)	(664.6)	(644.2)
Hhead educ	556.5**	2,121***	6,316***	7,447***
	(218.1)	(739.3)	(1,092)	(1,408)
Hhead marital	1,235*	-3,467**	-1,786	-2,070
	(697.6)	(1,378)	(1,843)	(2,297)
Hhead labourer	-628.1	2,813**	5,144***	5,553**
	(790.1)	(1,329)	(1,615)	(2,222)
Hhead farmer	-128.2	2,141*	2,901*	3,759*
	(693.2)	(1,241)	(1,530)	(1,958)
Hhead unemployed	-1,188	554.5	1,024	-4,661*
	(1,230)	(1,824)	(2,275)	(2,758)
Hhead gender	1,022	-136.1	7,888***	8,088
	(859.4)	(2,396)	(2,464)	(5,260)
Hhead age	-59.81***	655.1***	1,027***	1,069***
	(17.41)	(125.3)	(111.9)	(115.9)
Nb of migrants	-1,193***	157.4	-522.7	-3,546***
	(266.0)	(374.2)	(461.3)	(664.9)
HH size	2,464***	1,797***	2,591***	3,793***
	(181.5)	(380.0)	(401.0)	(486.6)
Nb of children	-1,866***	-1,313*	-2,663***	-3,927***
	(226.2)	(673.0)	(505.9)	(630.0)
Number of senior in HH	-1,273***	-23.61	-375.1	-1,216
	(233.4)	(810.6)	(707.2)	(846.7)
Observations	6,525	10,617	10,089	8,254
Number of HH	N/A	5,405	5,133	4,245

HH stands for 'household' and Hhead for 'household head'.
Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). Differences in sample size between regressions might arise from attrition (especially 2015). All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B9 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on health expenditure (using diff-in-diff strategy)

	(1)	(2)	(3)	(4)
		Health expenditure		
Health shock 2009	**2,388*** **			
	(228.8)			
Health shock 2009-2011		**1,379*** **		
		(305.9)		
Health shock 2009-2012			**605.2**	
			(378.1)	
Health shock 2009-2015				**600.6**
				(466.7)
House index	80.32	120.5*	88.47	411.3***
	(53.00)	(70.85)	(89.01)	(84.36)
Hhead educ	-21.99	-27.43	623.0***	465.4*
	(47.54)	(157.1)	(191.5)	(262.2)
Hhead marital	211.0	76.65	768.2**	-227.0
	(169.4)	(311.9)	(361.0)	(372.3)
Hhead labourer	247.0*	391.1*	393.9	224.2
	(139.3)	(218.1)	(249.6)	(297.0)
Hhead farmer	134.1	276.0	243.9	-80.87
	(118.9)	(198.0)	(245.4)	(283.8)
Hhead unemployed	619.7*	125.9	315.3	561.5
	(322.0)	(411.2)	(467.2)	(532.6)
Hhead gender	165.7	-850.1	1,620**	1,259**
	(227.1)	(710.5)	(641.6)	(565.8)
Hhead age	-1.528	9.314	70.15***	87.34***
	(4.588)	(16.60)	(16.42)	(12.95)
Nb of migrants	-24.38	-55.90	59.00	-212.9**
	(49.18)	(67.13)	(78.85)	(88.91)
HH size	121.9***	58.98	50.03	79.10
	(32.83)	(65.97)	(67.22)	(75.22)
Nb of children	-78.22	-20.65	-21.14	-93.07
	(48.87)	(95.35)	(93.12)	(90.83)
Number of senior in HH	271.1***	529.2***	598.0***	749.4***
	(63.08)	(130.6)	(148.3)	(134.3)
Observations	6,665	10,797	10,196	8,361
Number of HH		5,426	5,140	4,262

HH stands for 'household' and Hhead for 'household head'.
Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). Differences in sample size between regressions might arise from attrition (especially 2015). All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B10 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on food consumption (using diff-in-diff strategy)

	(1)	(2)	(3)	(4)
		Food consumption		
Health shock 2009	**112.6**			
	(170.3)			
Health shock 2009-2011		**-1,023*** **		
		(323.1)		
Health shock 2009-2012			**-1,173*** **	
			(297.3)	
Health shock 2009-2015				**-1,431*** **
				(361.5)
House index	481.4***	594.0***	622.7***	1,092***
	(82.46)	(98.21)	(105.2)	(97.17)
Hhead educ	-52.14	387.7*	1,246***	805.0***
	(64.90)	(217.1)	(226.2)	(229.9)
Hhead marital	534.4***	-699.8	-764.4*	-390.7
	(196.2)	(435.0)	(393.9)	(434.2)
Hhead labourer	-205.8	909.9***	310.7	664.0*
	(214.6)	(305.8)	(291.7)	(353.0)
Hhead farmer	319.2	681.2**	202.4	407.2
	(196.5)	(278.0)	(261.3)	(303.6)
Hhead unemployed	-678.9**	9.578	-130.9	-780.7
	(318.0)	(431.7)	(414.0)	(529.4)
Hhead gender	-305.9	359.4	685.3	274.3
	(226.9)	(641.0)	(660.5)	(790.9)
Hhead age	-2.100	159.6***	194.6***	124.8***
	(5.378)	(30.90)	(24.03)	(18.12)
Nb of migrants	-315.7***	-189.6**	-214.6**	-434.3***
	(76.56)	(95.24)	(95.05)	(105.8)
HH size	628.0***	467.8***	650.9***	466.0***
	(45.34)	(89.16)	(85.59)	(89.04)
Nb of children	-314.5***	-153.5	-267.3**	-136.7
	(57.10)	(143.9)	(105.3)	(102.5)
Number of senior in HH	-169.9**	-68.12	-35.64	-6.214
	(78.20)	(224.7)	(165.8)	(154.8)
Observations	6,545	10,596	10,106	8,231
Number of HH		5,420	5,139	4,254

HH stands for 'household' and Hhead for 'household head'.
Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). Differences in sample size between regressions might arise from attrition (especially 2015). All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B11 - Full regression tables for results presented in Chapter 5 on the long-term impact of health shocks (by type of shock) on NFNM (using diff-in-diff strategy)

	(1)	(2)	(3)	(4)
		NFNM consumption		
Health shock 2009	**338.7**			
	(562.1)			
Health shock 2009-2011		**-4,122*** **		
		(1,031)		
Health shock 2009-2012			**-4,325*** **	
			(1,045)	
Health shock 2009-2015				**-7,889*** **
				(1,557)
House index	3,686***	3,411***	4,446***	8,066***
	(285.0)	(357.5)	(445.4)	(472.3)
Hhead educ	631.8***	1,615***	3,598***	4,427***
	(189.3)	(567.6)	(775.8)	(999.0)
Hhead marital	816.8	-2,711**	-145.0	-4,497***
	(558.4)	(1,143)	(1,200)	(1,642)
Hhead labourer	-597.5	856.4	2,884**	5,656***
	(680.2)	(1,075)	(1,160)	(1,663)
Hhead farmer	-612.5	677.1	356.0	3,574**
	(597.9)	(976.1)	(1,090)	(1,516)
Hhead unemployed	-1,268	318.4	-1,695	-3,259
	(1,056)	(1,448)	(1,594)	(2,007)
Hhead gender	1,310*	500.8	7,044***	4,519
	(739.3)	(1,998)	(1,820)	(3,624)
Hhead age	-55.19***	441.9***	615.1***	666.7***
	(14.71)	(98.02)	(75.63)	(80.22)
Nb of migrants	-873.0***	378.0	-338.4	-1,944***
	(220.1)	(283.1)	(344.3)	(448.4)
HH size	1,787***	1,275***	1,677***	3,235***
	(157.5)	(301.4)	(318.3)	(369.1)
Nb of children	-1,502***	-1,506***	-2,220***	-3,647***
	(198.1)	(492.1)	(406.4)	(442.7)
Number of senior in HH	-1,434***	-307.1	-1,064*	-1,607***
	(190.2)	(568.8)	(600.8)	(538.1)
Observations	6570	10699	10112	8226
Number of HH		5414	5134	4252

HH stands for 'household' and Hhead for 'household head'.
Each column and line corresponds to a different regression (all estimates are obtained using first difference, with the exception of the first column which is obtained using cross-section). Differences in sample size between regressions might arise from attrition (especially 2015). All standard errors are clustered at the village level (in parentheses).
*** p<0.01, ** p<0.05, * p<0.1

Appendix B12 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on own work hours using a diff-in-diff strategy

	(1)	(2)	(3)	(4)
	All	Conditional on activity	Urban	Rural
		Own work hours		
Health shock	**-0.508*** **	**-0.354** **	**-0.267**	**-0.724*** **
	(0.127)	**(0.145)**	**(0.203)**	**(0.152)**
Age	0.0690	0.0603	0.0730	0.0421
	(0.0951)	(0.125)	(0.140)	(0.110)
Marital	0.0153	-0.109	-0.592*	0.637**
	(0.193)	(0.221)	(0.334)	(0.249)
Median consumption	0.142	0.144	0.134	0.0707
	(0.130)	(0.161)	(0.184)	(0.189)
Education	0.314*	0.0880	0.263	0.437**
	(0.159)	(0.271)	(0.236)	(0.209)
Nb of WA adults	-0.0986	-0.0843	0.0559	-0.144
	(0.0805)	(0.0966)	(0.109)	(0.105)
Nb of children	-0.0568	0.105	-0.153	0.00555
	(0.153)	(0.168)	(0.208)	(0.209)
Agricultural work	0.822**	-0.662**	0.402	1.162***
	(0.316)	(0.335)	(0.393)	(0.423)
Com consumption	1.46e-05	8.07e-06	-1.04e-05	5.23e-05*
	(1.40e-05)	(1.18e-05)	(1.54e-05)	(2.68e-05)
Electricity	-0.151	0.00498	-0.149	-0.119
	(0.104)	(0.118)	(0.159)	(0.150)
Toilet	0.00987	0.0152**	0.00816	0.00920
	(0.00718)	(0.00704)	(0.0116)	(0.00862)
Water Access	-0.0994	-0.281	-0.346	-0.0225
	(0.246)	(0.301)	(0.458)	(0.285)
Urban dummy	-0.00801	0.129		
	(0.148)	(0.143)		
Observations	16,442	10,087	7,242	9,200

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1

Appendix B13 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on own work hours using a diff-in-diff strategy *(continued)*

	(1)	(2)	(3)	(4)	(5)
	Men	Women	Household head	Age <40	Age 40-55
			Spousal work hours		
Health shock	**-0.395***	**-0.572*****	**-0.384***	**-0.319**	**-0.602*****
	(0.218)	**(0.164)**	**(0.204)**	**(0.217)**	**(0.158)**
Age	0.121	-0.0827	0.143	-0.328	0.114
	(0.090)	(0.208)	(0.100)	(0.241)	(0.093)
Marital	0.275	-0.0678	0.0539	-0.164	0.0535
	(0.357)	(0.226)	(0.322)	(0.385)	(0.195)
Median consumption	0.395*	-0.0241	0.406*	0.129	0.135
	(0.224)	(0.160)	(0.226)	(0.187)	(0.159)
Education	0.270	0.343*	0.389	0.294	0.181
	(0.211)	(0.206)	(0.307)	(0.181)	(0.267)
Nb of WA adults	-0.198	-0.0382	-0.204	0.194	-0.165*
	(0.123)	(0.096)	(0.132)	(0.145)	(0.099)
Nb of children	0.472**	-0.452**	0.351	-0.345	0.107
	(0.202)	(0.181)	(0.219)	(0.314)	(0.146)
Agricultural work	0.654	0.843**	0.822*	0.664	0.882**
	(0.425)	(0.360)	(0.463)	(0.428)	(0.399)
Com. consumption	1.14e-05	1.71e-05	1.04e-05	2.30e-05	1.16e-05
	(1.98e-05)	(1.60e-05)	(2.03e-05)	(1.93e-05)	(1.71e-05)
Electricity	-0.177	-0.149	-0.0807	-0.226	-0.100
	(0.172)	(0.118)	(0.176)	(0.165)	(0.150)
Toilet	0.022***	-0.000	0.022***	0.007	0.012
	(0.007)	(0.01)	(0.008)	(0.009)	(0.009)
Water Access	-0.056	-0.164	-0.014	0.011	-0.151
	(0.385)	(0.281)	(0.392)	(0.418)	(0.293)
Urban dummy	-0.048	0.039	0.237	-0.06	0.073
	(0.249)	(0.136)	(0.178)	(0.158)	(0.195)
Observations	7,660	8,782	6,420	8,210	8,283

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1
NB: those results are not conditional on employment (include zeros)

Appendix B14 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on inactivity (dummy variable) using a linear probability model

	(1)	(2)	(3)
	All	Urban	Rural
	Inactivity (dummy)		
Health shock	**0.0362*** **	**0.0387***	**0.0375****
	(0.013)	**(0.02)**	**(0.018)**
Age	-0.001	-0.02	0.013
	(0.013)	(0.015)	(0.010)
Marital	-0.024	-0.026	-0.012
	(0.027)	(0.045)	(0.034)
Median consumption	-0.015	-0.013	-0.015
	(0.014)	(0.0214)	(0.0190)
Education	-0.0385*	-0.032	-0.047*
	(0.019)	(0.022)	(0.028)
Nb of WA adults	-0.004	-0.005	-0.005
	(0.011)	(0.013)	(0.014)
Nb of children	0.007	0.014	0.003
	(0.016)	(0.020)	(0.020)
Agricultural work	-0.200***	-0.125***	-0.255***
	(0.038)	(0.043)	(0.055)
Com consumption	-1.34e-06	-1.84e-09	-3.51e-06
	(1.80e-06)	(1.97e-06)	(3.68e-06)
Electricity	0.030*	0.021	0.036
	(0.015)	(0.018)	(0.023)
Toilet	-0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.001)
Water Access	-0.009	-0.040	0.010
	(0.027)	(0.038)	(0.035)
Urban dummy	0.013		
	(0.018)		
Observations	16,442	10,087	7,242

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, linear probability, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1
NB: those results are not conditional on employment (include zeros)

Appendix B15 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on inactivity (dummy variable) using a linear probability model *(continued)*

	(1)	(2)	(3)	(4)	(5)
	Men	Women	Household head	Age <40	Age 40-55
			Inactive (dummy)		
Health shock	**0.023**	**0.039****	**0.025**	**0.021**	**0.043*****
	(0.020)	**(0.019)**	**(0.019)**	**(0.026)**	**(0.015)**
Age	-0.015	0.031	-0.014	0.037	-0.005
	(0.010)	(0.030)	(0.011)	(0.040)	(0.012)
Marital	-0.050	-0.009	-0.003	-0.048	-0.004
	(0.039)	(0.035)	(0.032)	(0.046)	(0.032)
Median consumption	-0.017	-0.013	-0.029	-0.012	-0.020
	(0.019)	(0.018)	(0.020)	(0.022)	(0.018)
Education	-0.015	-0.060**	-0.011	-0.051*	-0.014
	(0.027)	(0.023)	(0.034)	(0.027)	(0.026)
Nb of WA adults	0.005	-0.012	0.007	-0.016	-0.002
	(0.013)	(0.014)	(0.012)	(0.023)	(0.011)
Nb of children	-0.036**	0.040*	-0.020	0.051*	-0.017
	(0.016)	(0.021)	(0.015)	(0.030)	(0.016)
Agricultural work	-0.147***	-0.232***	-0.170***	-0.137***	-0.234***
	(0.048)	(0.044)	(0.047)	(0.051)	(0.041)
Com. consumption	-2.50e-06	-4.40e-07	-2.79e-06	-1.53e-06	-1.51e-06
	(2.32e-06)	(2.12e-06)	(2.31e-06)	(2.35e-06)	(1.95e-06)
Electricity	0.014	0.045***	0.015	0.042*	0.023
	(0.021)	(0.0168)	(0.020)	(0.024)	(0.018)
Toilet	-0.002	0.001	-0.002	-0.009	-0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Water Access	0.018	-0.019	0.022	0.001	-0.014
	(0.036)	(0.036)	(0.035)	(0.051)	(0.030)
Urban dummy	0.042*	-0.011	0.022	0.003	0.023
	(0.025)	(0.018)	(0.020)	(0.022)	(0.025)
Observations	7,660	8,782	6,420	8,210	8,283

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, linear probability, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1
NB: those results are not conditional on employment (include zeros)

Appendix B16 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on spousal work hours using a diff-in-diff strategy

	(1)	(2)	(3)	(4)
	All	Conditional on activity	Urban	Rural
		Spousal work hours		
Health shock	**0.237**	**0.078**	**0.075**	**0.389**
	(0.166)	**(0.179)**	**(0.246)**	**(0.246)**
Age	0.237***	0.232*	0.226***	0.310
	(0.086)	(0.124)	(0.051)	(0.238)
Median consumption	0.182	0.01	0.013	0.156
	(0.201)	(0.208)	(0.283)	(0.271)
Education	0.554*	0.411	1.024**	0.330
	(0.291)	(0.345)	(0.472)	(0.355)
Nb of WA adults	0.022	0.005	-0.007	0.013
	(0.125)	(0.137)	(0.177)	(0.171)
Nb of children	-0.075	0.009	-0.007	0.075
	(0.178)	(0.202)	(0.318)	(0.198)
Agricultural work	1.065**	-0.138	0.237	1.437*
	(0.520)	(0.367)	(0.520)	(0.741)
Com. consumption	5.14e-05***	3.38e-05**	1.55e-05	9.79e-05***
	(1.84e-05)	(1.65e-05)	-1.52E-05	(3.61e-05)
Electricity	-0.002	0.052	0.206	-0.164
	(0.156)	(0.168)	(0.192)	(0.235)
Toilet	-0.001	-0.007	-0.002	0.001
	(0.006)	(0.006)	(0.017)	(0.007)
Water Access	0.048	-0.075	-0.613	0.132
	(0.302)	(0.279)	(0.432)	(0.367)
Urban dummy	-0.180	0.250		
	(0.214)	(0.203)		
Observations	9,984	6,892	4,760	5,224

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1

Appendix B17 - Full regression tables for results presented in Chapter 6 on the impact of health shocks on spousal work hours using a diff-in-diff strategy *(continued)*

	(1)	(2)	(3)	(4)	(5)
	Men	Women	Household head	Age <40	Age 40-55
			Own work hours		
Health shock	**0.137**	**0.386***	**0.144**	**-0.162**	**0.402****
	(0.244)	**(0.218)**	**(0.251)**	**(0.346)**	**(0.187)**
Age	0.331***	-0.195	0.336**	0.275*	0.236**
	(0.114)	(0.143)	(0.129)	(0.141)	(0.101)
Median consumption	0.011	0.358	-0.045	0.121	0.201
	(0.268)	(0.254)	(0.269)	(0.317)	(0.216)
Education	0.753*	0.362	0.613	0.518	0.512
	(0.407)	(0.364)	(0.404)	(0.592)	(0.324)
Nb of WA adults	0.020	0.009	-0.053	0.028	0.067
	(0.177)	(0.161)	(0.174)	(0.329)	(0.126)
Nb of children	0.299	-0.503**	0.104	0.545*	-0.368*
	(0.288)	(0.239)	(0.294)	(0.302)	(0.205)
Agricultural work	1.564**	0.652	1.519**	0.018	1.549***
	(0.681)	(0.545)	(0.713)	(0.694)	(0.553)
Com. consumption	5.70e-05**	4.18e-05**	5.34e-05*	6.07e-05**	4.72e-05**
	(2.70e-05)	(1.90e-05)	(2.72e-05)	(2.51e-05)	(1.98e-05)
Electricity	-0.073	0.079	-0.062	0.223	-0.077
	(0.245)	(0.175)	(0.255)	(0.271)	(0.160)
Toilet	-0.003	0.009	0.003	0.005	-0.003
	(0.01)	(0.007)	(0.010)	(0.016)	(0.007)
Water Access	0.287	-0.241	0.229	0.581	-0.162
	(0.488)	(0.390)	(0.534)	(0.435)	(0.375)
Urban dummy	-0.215	-0.108	-0.205	-0.246	-0.177
	(0.312)	(0.254)	(0.324)	(0.274)	(0.245)
Observations	4,579	5,405	4,685	4,150	5,834

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1
NB: those results are not conditional on employment (include zeros)

Appendix B18 - Full regression tables for results presented in Chapter 6 on the impact of household health shocks on work hours using a diff-in-diff strategy

	(1)	(2)	(3)	(4)
	All	Conditional on activity	Urban	Rural
	Work hours – other household members			
Health shock	**0.055**	**0.009**	**0.149**	**-0.024**
	(0.119)	**(0.138)**	**(0.150)**	**(0.167)**
Age	0.003	-0.431***	-0.313	-0.038
	(0.092)	(0.109)	(0.234)	(0.122)
Marital	0.011	-0.248	0.036	-0.035
	(0.169)	(0.182)	(0.241)	(0.254)
Median consumption	-0.024	0.066	0.168	-0.141
	(0.125)	(0.152)	(0.143)	(0.172)
Education	0.023	-0.206	0.084	-0.065
	(0.09)	(0.180)	(0.101)	(0.145)
Nb of WA adults	0.091	0.069	0.152	0.016
	(0.085)	(0.110)	(0.103)	(0.115)
Nb of children	0.175	0.155	-0.132	0.348**
	(0.112)	(0.136)	(0.186)	(0.135)
Agricultural work	0.950***	0.826**	0.353	1.334***
	(0.348)	(0.376)	(0.323)	(0.447)
Com consumption	3.13e-05***	3.35e-05***	1.30e-05	5.15e-05**
	(1.16e-05)	(1.27e-05)	(9.54e-06)	(2.54e-05)
Electricity	-0.290***	-0.174	-0.332**	-0.306**
	(0.096)	(0.124)	(0.128)	(0.144)
Toilet	0.002	0.003	-0.008	0.004
	(0.008)	(0.008)	(0.011)	(0.010)
Water Access	-0.197	-0.270	-0.968***	0.157
	(0.209)	(0.282)	(0.278)	(0.270)
Urban dummy	0.022	-0.053		
	(0.110)	(0.143)		
Observations	15,031	10,888	6,960	8,071

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicate the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1

Appendix B19 - Full regression tables for results presented in Chapter 6 on the impact of household health shocks on work hours using a diff-in-diff strategy *(continued)*

	(1)	(2)	(3)	(4)	(5)
	Men	Women	Household head	Age <40	Age 40-55
	Work hours – other household members				
Health shock	**0.018**	**0.099**	**-0.008**	**-0.014**	**0.091**
	(0.166)	**(0.138)**	**(0.171)**	**(0.175)**	**(0.158)**
Age	0.004	0.001	-0.074	0.137	-0.157
	(0.121)	(0.085)	(0.120)	(0.107)	(0.106)
Marital	0.259	-0.215	0.012	-0.185	0.216
	(0.230)	(0.264)	(0.264)	(0.221)	(0.267)
Median consumption	0.034	-0.101	0.006	-0.015	-0.049
	(0.180)	(0.137)	(0.189)	(0.154)	(0.182)
Education	0.134	-0.108	0.599**	-0.155	0.539**
	(0.140)	(0.109)	(0.263)	(0.101)	(0.264)
Nb of WA adults	0.226**	0.004	0.107	0.232**	5.88e-05
	(0.109)	(0.094)	(0.133)	(0.107)	(0.118)
Nb of children	0.413**	-0.085	0.381**	0.242*	0.058
	(0.164)	(0.129)	(0.192)	(0.135)	(0.149)
Agricultural work	1.067**	0.848***	1.553***	0.577	1.582***
	(0.464)	(0.317)	(0.519)	(0.376)	(0.466)
Com. consumption	2.15e-05	3.91e-05***	1.93e-05	1.90e-05	5.24e-05***
	(1.43e-05)	(1.28e-05)	(1.69e-05)	(1.23e-05)	(1.70e-05)
Electricity	-0.283**	-0.286**	-0.176	-0.366***	-0.175
	(0.128)	(0.120)	(0.146)	(0.120)	(0.156)
Toilet	0.006	-0.002	0.003	0.005	-0.001
	(0.012)	(0.007)	(0.010)	(0.011)	(0.008)
Water Access	-0.345	-0.051	0.029	-0.094	-0.425
	(0.325)	(0.213)	(0.367)	(0.250)	(0.300)
Urban dummy	0.061	0.022	0.004	0.102	-0.137
	(0.157)	(0.112)	(0.157)	(0.141)	(0.128)
Observations	7,006	8,025	5,809	7,487	7,544

WA adults corresponds to 'working age adults' (age 16-55)
Com. Consumption refers to the average consumption in the village
Median consumption is a dummy variable that corresponds to whether the household is below the median of consumption in its village
Each column corresponds to a different regression (first difference, IPTW to control for differences in characteristics based on propensity score for own health shock) – labels on the top of the column indicates the sample on which each regression is performed. Standard errors are clustered at the village-level.
*** p<0.01, ** p<0.05, * p<0.1
NB: those results are not conditional on employment (include zeros)

Appendix B20 - Full regression tables for results presented in Chapter 7 on the impact of health shocks the decision to return (own decision) (fixed effects logit)

	(1) All migrants	(2) Male	(3) Female	(4) HH head	(5) age<40	(6) Age 40-55
			Return migration			
Health shock	0.334**	0.304**	0.068	0.221	0.199	0.318
	(0.142)	-0.128	(0.361)	(0.184)	(0.187)	(0.239)
Age	0.757***	0.662***	1.081***	0.409	1.201***	0.935
	(0.147)	(0.173)	(0.295)	(0.334)	(0.261)	(0.837)
Age squared	-0.010***	-0.008***	-0.015***	-0.005	-0.018***	-0.011
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.008)
Marital status	0.275	0.108	0.832**	0.572	0.341*	-0.413
	(0.168)	(0.193)	(0.360)	(0.451)	(0.184)	(0.494)
Primary school	-0.342	-0.410	-0.231	-0.764**	-0.119	-0.587
	(0.236)	(0.287)	(0.429)	(0.377)	(0.301)	(0.400)
Middle School	-0.293	-0.468	0.078	-0.702	-0.054	-0.687
	(0.273)	(0.327)	(0.524)	(0.475)	(0.326)	(0.566)
High school/univ	0.154	-0.020	0.756	-0.903	0.585	-1.495
	(0.381)	(0.438)	(0.806)	(0.863)	(0.446)	(0.975)
HH size	0.009	0.022	-0.034	-0.051	0.094	-0.092
	(0.061)	(0.067)	(0.147)	(0.084)	(0.089)	(0.088)
Nb of seniors	-0.056	-0.078	0.061	-0.057	-0.094	-0.014
	(0.061)	(0.067)	(0.158)	(0.089)	(0.075)	(0.111)
Nb of children	-0.009	-0.046	0.117	0.092	-0.068	0.105
	(0.101)	(0.114)	(0.222)	(0.140)	(0.136)	(0.170)
Com consumption	-4.85e-06	-4.75e-06	-6.03e-06	-7.77e-08	-1.08e-05	4.99e-06
	(5.02e-06)	(5.66e-06)	(1.13e-05)	(7.33e-06)	(6.63e-06)	(8.22e-06)
House index	0.0911*	0.0413	0.274**	0.0528	0.117	0.0726
	(0.0551)	(0.0622)	(0.121)	(0.0759)	(0.0709)	(0.0919)
Observations	3,750	2,844	906	1,802	2,328	1,334
Nb of individuals	1,875	1,422	453	901	1,164	667

'Univ' refers to 'University' (High school/Univ is a dummy variable for whether the respondent has gone to high school or more).
'Com consumption' refers to 'community consumption'
Full regression table obtained by fixed effects logit models. The dependent variable is 'return migration' and the predictor of interest is health shocks. Each column corresponds to a separate estimate, first column corresponding to all migrants (1), then split by gender (2-3), presented for household head only (4), and by age group (5-6)
*** p<0.01, ** p<0.05, * p<0.1

Appendix B21 - Full regression tables for results presented in Chapter 7 on the impact of health shocks the decision to return (fixed effects logit) (own decision) *(continued)*

	(1)	(2)	(3)	(4)
	<9 months migration	>=9 months migration	admission < or = 7 days	admission over 7 days
	Return migration			
Health shock	**0.336**	**0.221**		
	(0.220)	**(0.310)**		
Short admission shock			**0.284**	
			(0.189)	
Long admission shock				**0.393****
				(0.176)
Age	0.696***	1.108***	0.754***	0.755***
	(0.190)	(0.229)	(0.146)	(0.147)
Age squared	-0.007***	-0.013***	-0.009***	-0.009***
	(0.002)	(0.003)	(0.002)	(0.002)
Marital status	0.350	0.190	0.271	0.277*
	(0.227)	(0.238)	(0.168)	(0.168)
Primary school	-0.419	-0.114	-0.337	-0.338
	(0.274)	(0.395)	(0.236)	(0.236)
Middle School	-0.238	-0.084	-0.293	-0.288
	(0.326)	(0.444)	(0.273)	(0.273)
High school/univ	-0.0609	0.859	0.155	0.160
	(0.460)	(0.669)	(0.380)	(0.381)
HH size	-0.071	0.147	0.007	0.007
	(0.074)	(0.107)	(0.061)	(0.061)
Nb of seniors	-0.084	0.070	-0.055	-0.055
	(0.073)	(0.116)	(0.061)	(0.061)
Nb of children	0.060	-0.021	-0.009	-0.010
	(0.121)	(0.178)	(0.101)	(0.101)
Com consumption	-1.72e-06	-9.69e-06	-4.90e-06	-4.70e-06
	(6.32e-06)	(7.80e-06)	(5.02e-06)	(5.02e-06)
House index	0.075	0.206**	0.091*	0.092*
	(0.064)	(0.096)	(0.055)	(0.055)
Observations	2,572	1,430	3,750	3,750
Nb of individuals	1,286	715	1,875	1,875

'Univ' refers to 'University' (High school/Univ is a dummy variable for whether the respondent has gone to high school or more).
'Com consumption' refers to 'community consumption'
Full regression table obtained by fixed effects logit models. The dependent variable is 'return migration' and the predictor of interest is health shocks. Each column corresponds to a separate estimate, first and second column corresponding to groups based on migration length, and third and fourth column to severity of the shock as proxied by length of stay.
*** p<0.01, ** p<0.05, * p<0.1

Appendix B22 - Full regression tables for results presented in Chapter 7 on the impact of household health shocks the decision to return (other household members) (fixed effects logit)

	(1) All migrants	(2) Male	(3) Female	(4) HH head	(5) age<40	(6) Age 40-55
	Return migration (other household member health shock)					
Health shock (other HH members)	0.133	0.153	0.0461	0.121	0.0872	0.281
	(0.104)	(0.120)	(0.217)	(0.158)	(0.129)	(0.190)
Age	0.763***	0.653***	1.097***	0.373	1.359***	0.987
	(0.152)	(0.180)	(0.308)	(0.356)	(0.270)	(0.879)
Age squared	-0.009***	-0.008***	-0.015***	-0.005	-0.021***	-0.011
	(0.002)	(0.002)	(0.004)	(0.004)	(0.005)	(0.009)
Marital status	0.224	0.019	0.868**	0.855*	0.234	-0.181
	(0.173)	(0.201)	(0.364)	(0.497)	(0.188)	(0.516)
Primary school	-0.423*	-0.421	-0.460	-0.839**	-0.167	-0.688*
	(0.245)	(0.294)	(0.453)	(0.391)	(0.313)	(0.414)
Middle School	-0.368	-0.481	-0.122	-0.881*	-0.109	-0.807
	(0.279)	(0.332)	(0.536)	(0.489)	(0.335)	(0.573)
High school/univ	0.186	0.020	0.898	-0.858	0.603	-1.414
	(0.392)	(0.446)	(0.865)	(0.905)	(0.459)	(0.995)
HH size	0.032	0.063	-0.083	-0.006	0.130	-0.086
	(0.063)	(0.070)	(0.152)	(0.088)	(0.092)	(0.091)
Nb of seniors	-0.076	-0.111	0.086	-0.119	-0.121	-0.013
	(0.065)	(0.071)	(0.168)	(0.100)	(0.080)	(0.113)
Nb of children	-0.077	-0.116	0.077	0.020	-0.152	0.061
	(0.104)	(0.117)	(0.232)	(0.145)	(0.140)	(0.176)
Com consumption	-5.17e-06	-4.22e-06	-1.02e-05	8.81e-07	-1.17e-05*	5.72e-06
	(5.21e-06)	(5.88e-06)	(1.17e-05)	(7.76e-06)	(6.81e-06)	(8.70e-06)
House index	0.113**	0.065	0.295**	0.078	0.165**	0.064
	(0.057)	(0.065)	(0.125)	(0.080)	(0.073)	(0.095)
Observations	3,536	2,678	858	1,662	2,220	1,232
Nb of individuals	1,768	1,339	429	831	1,110	616

'Univ' refers to 'University' (High school/Univ is a dummy variable for whether the respondent has gone to high school or more).
'Com consumption' refers to 'community consumption'
Full regression table obtained by fixed effects logit models. The dependent variable is 'return migration' and the predictor of interest is health shocks. Each column corresponds to a separate estimate, first column corresponding to all migrants (1), then split by gender (2-3), presented for household head only (4), and by age group (5-6)
*** p<0.01, ** p<0.05, * p<0.1