UNIVERSITA' CATTOLICA DEL SACRO CUORE
MILANO

Dottorato di Ricerca in Economia e Finanza
della Pubblica Amministrazione

Ciclo XXIV

S.S.D: SECS-P/03 AND SECS-P/05

Health Care System Reforms and Medical Poverty Trap in Rural China

Tesi di Dottorato di: Wei HAN
Matricola: 3703874

Anno Accademico 2010/11

UNIVERSITA' CATTOLICA DEL SACRO CUORE
MILANO

DEFAP
Dottorato di Ricerca in Economia e Finanza della Pubblica Amministrazione

Ciclo XXIV

S.S.D: SECS-P/03 AND SECS-P/05

Health Care System Reforms and Medical Poverty Trap in Rural China

Coordinatore: Ch.mo Prof. Massimo BORDIGNON

Tesi di Dottorato di: Wei HAN
Matricola: 3703874

Anno Accademico 2010/11

Table of contents

Chapter one: Health care system reform in developing countries: Why some of them do not improve financial protection for the poor?

Table of contents (Continuing)

Chapter two: Catastrophic health expenditure and well-being in a multidimensional poverty framework

Table of contents (Continuing)

Chapter three: Evaluate the impact of provider payment intervention on out-of-pocket health expenditure: Evidence from a randomized social experiment in rural China

Health care system reform in developing countries: Why some of them do not improve financial protection for the poor?

Abstract

Most empirical studies have shown that health care system reform in China does not work as expected. By comparing the case of China with the cases of Mexico and Vietnam, we try to find the explanation from two aspects: policy design and evaluation methodology. It is possible to comment on the single policy design to some extent, whereas it is actually implausible to judge the policy package as a whole since governments always face various constraints. However, it is highly possible that the non-positive empirical evidence in the case of China is caused by the evaluation error rather than the policy design itself. We further elaborate the evaluation issues from the choice of outcome variables, data and estimation strategies.

***Keywords*:** health system reform, developing countries, policy design and evaluation

1. Introduction

The World Health Report (WHO 2000) pointed out three fundamental objectives of health systems, which are improving the health of the population they serve, responding to people's expectations as well as providing financial protection against the costs due to illness. In addition, it laid stress on that the poor needed financial protection as much as or more than the well-off, since even small absolute risks may give birth to catastrophic consequences for them. In recent years, many lower to middle income countries, including China, Vietnam, India, Colombia, Mexico, Thailand, Kenya, Ghana, Zambia, etc. are reforming their health systems for promoting universal access of health care, improving equity of health, quality of health service, and fairness of financing. Most of the reforms appeal to social health insurance as the main approach to improve health care systems so as to protect the poor.

The first debate around whether the social health insurance system should be established or not can date back to more than 45 year ago. Arrow (1963) argued that the extension of insurance coverage to the rational and risk averse uninsured was overwhelming if fair premium was charged. Given the premium of social health insurance is relatively low as it is usually subsidized by public finance, it should be further promoted by the government, especially to the uninsured poor. However, according to the comment of Pauly (1968), this result may not hold under the influence of moral hazard[1] since the insured patients respond to the reduced medical price by taking more health care that they may not need at the full price. That is to say, the insured incur a welfare loss by consuming additional health care.

Empirical studies show that some of the reforms do provide financial protection but some do not. Most of the studies do not report the positive impact in the term of China. This chapter aims at shedding light on why the reform in China seems not to work as it supposed to be by means of reviewing the existing literature of health system reform in two developing countries –Mexico and Vietnam- and then comparing them with the completed studies in China. The cross-country comparison is conducted from the following perspectives: (1) the policy design and implementation; and (2) evaluation methodology. The former one is further elaborated from background differences, identification of beneficiary, design of benefit package and theoretical foundation. The latter one covers the choice of outcome variable, data quality and evaluation methodology.

[1] Moral hazard can be defined as any change in behaviour that is due to becoming insured. For instance, the insured may take fewer precautions to avoid becoming ill, which is ex-ante moral hazard; the insured patients consume more medical care, which is ex-post moral hazard.

Mexico and Vietnam are chosen as the representatives of the health care system reforms[2]. Mexico is chosen as one of our comparison due to the following reasons. On the one hand, although there is room of improvement and innovation, Mexico could be regarded as one of the most successful cases in the health care system reform in developing countries, which could be of help to highlight the blank and potential problems hiding in the case of China. On the other hand, in terms of the social health insurance, the key component of health care system reform, it is promoting the voluntary universal social health insurance that is the same as in China. However, Vietnam, as the neighbor of China, has more common ground in the demographic factors. Moreover, in contrary to the cases of China and Mexico, it is implementing the compulsory targeting social health insurance program to the identified poor, which provides an opportunity to compare the effectiveness of allocating public finance in different ways.

The paper is organised as follows. Section 2 describes health care system reform in Mexico and Vietnam and summarizes their evaluation results. Section 3 focuses on the reform in China. Section 4 and Section 5 tries to answer the research question from the perspective of policy design and empirical evaluation, respectively. Section 6 concludes.

2. Health care system reform in developing countries: Mexico and Vietnam

Our discussion concentrates on the poor or the disadvantaged groups in the developing countries. For each country, general information of health system reform will be provided, including the key timeline, the aims and the main components of the reform. In terms of the last one, our discussion will focus on demand side -social health insurance- and supply side –provider payment system. Given the main innovation of most health system reforms lies on the social health insurance, it will be further elaborated as follows: (1) the configuration of the social health insurance if it is relevant to our discussion; (2) the detailed introduction about the specific social health insurance targeting the poor or the disadvantaged groups, including who or how to identify the eligible population, who are the main contributors and what is the benefit package; and (3) the corresponding stewardship reform. Last but not the least, the evaluation results will be summarized.

2.1. Mexico

According to World Health Report, one of the major problems in Mexico is catastrophic health spending (WHO 2000). More than half of its citizens were potentially exposed to the financial risk due to health care (Frenk, et al., 2003), which was caused by the social inequalities in the process of

[2] Apparently, India could play as role as representative too. Nevertheless, we did not choose it due to the reasons as follows. First, they started a serious effort on the social health insurance for the poor four years ago. Rajiv Aarogyasri Scheme (RAS), the first of this class targeting below-the-poverty-line population of Andhra Pradesh was introduced in 2007. Some other state governments are still in the process of introducing insurance schemes targeting poor households. That means we cannot get enough evaluation results by now. Second, each state is promoting their own community health insurance with a large variety. Generally speaking, it is feasible and meaningful for across-state comparison but not the international comparison. For the details, please see Reddy et al. (2011).

development but also deteriorated the social inequalities. Since 2004, Mexico has been implementing a major health system reform and established System of Social Protection in Health (SSPH) to overcome the health-related problems of the uninsured, most of which are the poor, based on the idea that health care is a social right rather than a commodity or a privilege (Frenk et al. 2006). A empirical study conducted in 2002 shows the excessive health-related spending for the poorest rural families in Mexico, most of which were attributed to outpatient care and medication (Galárraga et al. 2010)

The main innovation of the reform is called Seguro Popular, which aims at providing financial protection associated with illness to those who are not covered by any other public health schemes[3]. It is worth emphasizing that Seguro Popular has thoroughly broken the barriers of participating public health schemes due to the poverty or labour market status, i.e. all individuals who do not benefit from social security are eligible no matter they are self-employed, unemployed, or even out of the work force. It is a subsidised voluntary insurance that offers free access at the point of delivery to an explicit set of health care services, which is divided into two parts: an essential package of primary and secondary-level care and a package of high-cost tertiary-level care associated with catastrophic health expenditure. There are three sources of contributions: government, states and families. Public finance contributes the majority of the funds while families only pay a small premium through a progressive, means-tested sliding scale. The poorest 20% of families are exempt from any contribution. Moreover, in addition to the necessary stewardship reform on the organizational structure, Mexico has built up a comprehensive monitoring and evaluation system to support the policy design and implementation of Seguro Popular, such as publish benchmark report on a year base, conduct special-designed longitudinal survey from 2005 to evaluate the progress. The problems related to provider payment is not relevant since all the health interventions and drugs defined in the benefit package are free to access, which coverers 266 unique health interventions, 312 medicines in 2006. In fact, before the reform, the uninsured population had access to health facilities run by state but needed to pay user fee, which brought about the high proportion of out-of-pocket expenditure. Moreover, the uninsured suffered an additional out-of-pocket expenditure due to the shortages of drugs as a result of budgetary limitations (Frenk et al., 2006). Seguro Popular is progressing on covering more people, more interventions, and more conditions, with better quality.

Knaul et al. (2006) presented 15-year trends on the evolution of catastrophic and impoverishing health expenditure. The time trends and the econometric analysis suggest that the reduction in out-of-pocket health expenditure and catastrophic health expenditure by households is related to the expansion of Seguro Popular, although no causality conclusion can be drew due to the data availability. Moreover, they found that some key components of Seguro Popular -insuring the poorest quintiles, covering medications and ambulatory care, and including a package of catastrophic expenditures- are effective

[3] There are two social health insurance schemes for the employees working in the private sector –IMSS- and public sector –ISSSTE. Federal government contributes partly and their employers play a role as co-responsible contributors.

strategies to reduce catastrophic and impoverishing health expenditure. Galárraga et al. (2008), the first paper reporting the causal effect using non-experimental data, analysed the impact of Seguro Popular on the incidence of catastrophic health expenditure and out-of-pocket health expenditure among poor households using three different data sources. The bivariate probit models were applied to take the endogeneity of insurance selection choices into account by means of instrumental variables. The study shows that there was a statistically significant effect on the reduction of households' expenditures on medicines and outpatient care regardless the data sources. However, the effect on catastrophic health expenditure differed with the data resources. A matched-pair cluster-randomised experiment carried out by King et al. (2009) confirmed that Seguro Popular had been successful in reducing overall catastrophic and out-of-pocket expenditure, especially in the poorest individuals; however, they reported no effects on medication spending, health outcomes, or utilisation, which is contrary to other observational studies and possibly due to the short duration of treatment (10 months).

2.2. Vietnam

The situation in Vietnam is similar to that in other developing countries in Asia in the sense that out-of-pocket expenditure is the dominant source of health financing (WHO 2006). Moreover, as one of the low-income countries that are aspiring to universal health insurance coverage, the experience of Vietnam can provide more lessons to China, especially to the rural area of China.

Vietnam's health care system reform focuses on the social insurance reform. Compared with other developing countries, Vietnam started to develop social health insurance much earlier. The process of introducing social health insurance was initiated in 1992. At the end of 2002, Health Care Fund for the Poor (HCFP) – a compulsory social health insurance program targeting all poor households and selected disadvantaged groups- was formally set up. By 2006, the program had been extended to around 60% of those eligible (Wagstaff 2010). The eligible population includes the poor and some socially protected population groups including people of merit, the elderly, and war dioxin victims[4]. All the eligible population can be clearly identified except for the so-called poor, which, in practice, are identified by existing lists of other government programs as well as household surveys (Björn Ekman et al. 2008). It is worth noting that self-employed, informal sector workers, dependents of Compulsory Health Insurance members cannot benefit from this program, neither from Compulsory Health Insurance. They are classified as the potential participants of Voluntary Health Insurance. In terms of the contribution, the central government and the provincial governments contribute 75% and 25%, respectively. In other words, there is no need of contribution from eligible population. On the contrary, Voluntary Health Insurance is funded by private premium contributions based on ability to pay. The Vietnamese health insurance system provides the insured with a relatively broad – covering

[4] Initially implemented as a separate social program, HCFP was rolled into the national compulsory health insurance (CHI) scheme in July of 2009. The current national health insurance system consists of two parts, compulsory health insurance (CHI) and voluntary health insurance (VHI).

outpatient as well as inpatient services at all health care levels, laboratory exams, x-ray, and so on- but unspecified benefit package, which is identical to all the insurance programs (Long 2007; Ekman et al. 2008). Fee-for-service (FFS) is used for both outpatient care and inpatient care, which may bring about the supply-induced demand of health care in the context of the social insurance, i.e. give inappropriate incentive to the physicians to over-treat the patients. Although some alternative provider payment methods have been suggested, such as capitation, it has not been largely extended in practice. In 2003, Vietnam Social Security (VSS) was founded to administrate all social insurance programmes, including collecting premiums, issuing health insurance cards and reimbursing service providers.

There are limited literatures evaluating the impact of HCFP. By comparing households in the program with ones outside it and employing propensity score matching on a trimmed cross-sectional data, Wagstaff (2007) found that HCFP may be well targeted on the Vietnam's poor since it had reduced the risk of catastrophic out-of-pocket spending but no perceptible impact on average out-of-pocket spending. However, Axelson et al. (2009) applied propensity score matching with both single differences and double differences to pre- intervention (2002) and post-intervention data (2004) from Vietnam Household Living Standards Surveys. They reported a small, positive impact on health care utilization but a strong negative impact on out-of-pocket health expenditure. As the time goes by, more and more data are collected and released. Wagstaff (2010) applied triple differencing with matching to estimate the impacts of HCFP, the results of which is robust to the biases due to unobserved heterogeneity. Two rounds prior to the program's implementation (2002 and 2004) and one round after (2006) were employed. They concluded that HCFP has considerably positive impact on reducing out-of-pocket health expenditure but not the health care utilization, which is inconsistent with Wagstaff (2007).

3. Health care system reform in developing countries: China

In 2003, the Chinese government initiated the New Cooperative Medical Scheme (NCMS), a co-payment voluntary insurance system subsidized by the central government and provincial governments. One of the primary goals of NCMS is to prevent its rural population from being impoverished by catastrophic health expenditure.

All the rural residents are eligible for NCMS. Given the voluntary nature of NCMS, the participation is required at the household level in order to address the problem of adverse selection. According to statistical communique on the 2009 national development of the medical and health service provided by Ministry of Health of China, by the end of 2009, NCMS has already covered more than 0.83 billion people, which is 94.0% of the target population. 94.4 billion RMB (approximately 10 billion euros)

was financed, around 75%[5] of which was subsidized by the central and local governments. Provincial and county governments have considerable discretion on how NCMS should be designed and implemented following the guidelines from central government, which brings about considerable heterogeneity in the benefit package across counties and areas. Generally speaking, due to limited financing, benefit package is not so generous, especially for outpatient care. Some services are not covered or covered only partially, deductibles are high, ceilings are low, and coinsurance rates are high Wagstaff et al. (2009). However, all counties cover inpatient care. In 2009, the Chinese government announced that it would spend an additional 125 billion USD over the next three years to further improve health care system. One of the specific areas is on NCMS (Yip & Hsiao 2011).

The existing literatures generally fall into three categories: statistic report without exploring the causal effect; causal effect analysis using econometric techniques and comparison with some other experimental program.

From a stratified cluster sample of rural households, Sun et al. (2009) refined a subsample which only included households had suffered catastrophic health expenditure during 2004 to measure the impact of the NCMS by counterfactual analysis, i.e. comparing catastrophic payments before and after NCMS reimbursements by the catastrophic headcount and catastrophic payment gap. It was concluded that both indicators were reduced after reimbursement; however, the majority of catastrophic households remained catastrophic. In addition, Zhang et al. (2010) took a random sample of NCMS enrolees who had obtained reimbursement to apply counterfactual analysis. They found that NCMS could reduce catastrophic health expenditure partially and the poor could be protected more than the non-poor. In these two studies, both direct medical expenditure and non-medical expenditure were considered as health expenditure. Nevertheless, the significant shortcoming of this kind of counterfactual analysis is that it did not consider the change of patients' health care seeking behaviour and providers' demand inducing behaviour after the implementation of a health insurance scheme. Sun et al. (2009) explained why they chose a counterfactual analysis rather than an 'idea' controlled study. Furthermore, the impact of NCMS on financial protection was also examined according to the real reimbursement rate, Yi et al. (2009) reported that NCMS did not meet the goals of helping rural residents deal with financial risks due to catastrophic illness.

The most important and popular category of literatures is that evaluating and appraising health care policy by means of causal effect analysis. Advanced microeconometric techniques such as difference-in-difference, instrumental variable, propensity score matching, fixed effect and random effect models, multivariate linear regression are put into use. Wagstaff et al. (2007, 2009) combined difference-in-difference with propensity score matching to estimate the impact of NCMS. A probit model with

[5] This is an approximate figure since the actual one is not yet available. Moreover, the central government only regulates the minimum standard. The detailed schemes are within the discretion of the local governments based on their actual situation.

dummy outcome variable is used to derive propensity score. They pointed out that out-of-pocket spending on health care did not appear to be reduced by NCMS, which may be attributed to the increase of the likelihood of people seeking medical care as well as the volume of care provided after extending NCMS. In contrast, by employing the difference-in-difference framework and a parametric regression counterpart, i.e. tobit model, Wu (2010) found that NCMS helped reducing total out-of-pocket expenditure for the participants, however, it seemed that most of the decreases come from the change in utilization instead of price effect. Wagstaff et al. (2007, 2009) admitted the limitation of their research as follows: the bias due to a lack of complete balancing on observable still existed; some unobserved time-varying heterogeneity likely remained; the sample of NCMS counties was not a random sample of the pilots and non-random program placement. Wu (2010) was confident to her findings since a much more informative data set was used comparing with that of Wagstaff et al. (2007, 2009), and it allowed the counterfactual time trends of the outcomes for the treated to differ from those for the controlled group. For the sake of correcting the potential selection bias, Lei & Lin (2009) applied individual fixed effect model, instrumental variable and propensity score matching with difference-in-difference estimation to three-wave panel data. There was no evidence that NCMS participation could relieve financial burden, as measured by out-of-pocket expenditures among patients[6]. Besides Wu (2010), Babiarz et al. (2010) showed a positive evaluation result. They estimated the effect of individual NCMS policy features using multivariate linear regressions that controlled for a set of clinic and individual attributes as well as village and year effects and found that both out-of-pocket medical spending and catastrophic spending fell with NCMS participation. Moreover, robustness was checked by propensity score matching, which yielded the same pattern of results with slightly larger magnitudes.

Some social experiments are conducted during the past few year in order to detect potential problems and explore the possible improvement of NCMS. A social experiment called Rural Mutual Health Care (RMHC) was conducted in 2006. It provided first dollar coverage for primary care, hospital services and drugs with a similar premium but a lower ceiling. Moreover, it changed the provider payment for village doctors from fee-for-service to salary plus performance based bonus and introduced bulk purchasing for drugs. Yip & Hsiao (2009) compared NCMS with RMHC using static simulation model. The results indicated that RMHC was more effective at reducing medical impoverishment than NCMS. However, their study did not take into account the behavioural responses of patients and providers under the different benefit package designs. Besides, data were based on a population that voluntarily selected to enrol in RMHC, which means their health expenditure distribution may be higher than that of the general population. World Bank financed a

[6] They instrumented for individual NCMS enrolment status using county NCMS enrolment status; Difference-in-difference propensity score matching compared differences before and after participating NCMS. For individuals enrolled in 2004 'before' was the situation of 2000. For those who started participating in 2006, 'before' was that of 2000 and 2004. The control group included individuals who did not participate in any of these three waves.

health system reform pilot in Gansu (China), which combined supply-side interventions aimed at improving the effectiveness and quality of care with demand-side interventions aimed at expanding health insurance and providing financial support to the poor. Wagstaff & Yu (2007) estimated the impacts of this health reform pilot by means of combining differences-in-differences with propensity score matching. They reported the pilot had a positive impact on reducing out-of-pocket health expenditure, and the incidence of catastrophic health expenditure and impoverishment through health expenses.

In the previous two sections, we have described the health care system reform in three developing countries –Mexico, Vietnam and China- with the special attention to their impacts on the poor, which aims at providing key background information and general results of the reforms. In the following two sections, we attempt to figure out why the health care system reform in China does not work as expected by comparing the cases of Mexico and Vietnam with the case of China. Our discussion will focus on two aspects, which are policy design and evaluation methodology.

4. Policy problem

Policy could differ in many aspects. We focus on the following: (1) the background differences such as social development, poverty level and population health; (2) the design of health care system reform, including the intervention policy on the demand side as well as supply side; and policy implementation and monitoring; (3) narrowing to the demand side social insurance reform, the target population, the design of benefit package and the theoretical foundation of social health insurance.

4.1. Background differences

Background differences among countries could be one the most important attributions of policy differences. In other words, applying the successful experience of other countries without taking the background factors into account may not bring about the expected results. Table 1 compares some relevant figures about social development, poverty level and population health of China, Mexico and Vietnam. In terms of social development, for average, Mexico is much better off compared with China and Vietnam. GDP per capita of Mexico is nearly twofold as that of China and even more than fourfold as that of Vietnam. Also, Mexico is at a high urbanization level while China and Vietnam are still on the halfway. However, Mexico has the most serious income inequality in the sense that the poorest 10% households consume 1.5% of total consumption while the richest 10% consume more than 40%. As far as the population health is concerned, China is facing a serious ageing problem due to the implementation of one-child policy, while Mexico is challenged by obesity problem. In 2009, Mexico's health expenditure accounted for more than 13% of GDP, which ranked the third in the world and was threefold as that of China. Note that, Mexico had a relatively high density of

physicians but very low density of hospital beds, although these data were collected from 2004. This may explain one of the master plans of health system reform in Mexico is about the investment in infrastructure, i.e. expanding the availability and capacity of health care facilities. In summary, Table 1 implies that these three countries have different starting points, which is of importance to the design of policy. For instance, if the poor are extremely poor, more resources are required to make a significant improvement. All in all, the design of health care system should reflect the characteristics of the society. Although Mexico has got great achievement in the reform of health care system, we should be cautious to apply their experience.

Table 1: Social development, poverty level and population health[7]

	China	Mexico	Vietnam
Population (million)	1336.72 (1)[8]	113.72 (11)	90.55 (14)
GDP per capita (PPP, $)	7600 (125)	13900 (85)	3100 (166)
Urbanization	47.0%	78.0%	30.0%
Labour force - by occupation			
- agriculture	38.1%	13.7%	53.9%
- industry	27.8%	23.4%	20.3%
- services	34.1%	62.9%	25.8%
Household consumption by percentage share			
- lowest 10%	3.5%	1.5%	3.2%
- highest 10%	15.0%	41.4%	30.2%
Age structure			
- 0-14	17.6%	28.2%	25.2%
- 15-64	73.6%	65.2%	69.3%
- >64	8.9%	6.6%	5.5%
- Median age	35.50	27.10	27.80
Birth rate (per 1000 population)	12.29 (159)	19.13 (102)	17.07 (118)
Death rate (per 1000 population)	7.03 (131)	4.86 (191)	5.96 (163)
Life expectancy	74.68 (95)	76.47 (72)	72.18 (129)
Obesity - adult prevalence rate	2.9% (66)	23.6% (13)	0.5% (70)
Health expenditure[9] (% of 2009 GDP)	4.6 (148)	13.8 (3)	7.2 (72)
Hospital bed density (per 1000 population)[10]	4.06 (48)	1.6 (118)	2.87 (80)
Physicians density (per 1000 population)	1.415 (86)	2.893 (42)	1.224 (90)
Improved drinking water source			
- urban	98%	96%	99%
- rural	82%	87%	92%
- total	89%	94%	94%
Improved sanitation facility access			
- urban	58%	90%	94%
- rural	52%	68%	67%
- total	55%	85%	75%

4.2. The design of health care system reform

Most of the health care system reforms in the developing countries focus on the demand side intervention, the main component of which is the introduction or extension of social health insurance. The case of Mexico, Vietnam and China all fall into this category. It is worth noting that, in terms of

[7] All the figures in this table are collected from The World Factbook, available from https://www.cia.gov/library/publications/the-world-factbook/

[8] Figures in the parenthesis are the corresponding rankings in the world.

[9] Health expenditures are broadly defined as activities performed either by institutions or individuals through the application of medical, paramedical, and/or nursing knowledge and technology, the primary purpose of which is to promote, restore, or maintain health.

[10] For the hospital bed density and physician density, figures are lack of comparison as they were collected in different years: 2009 (China), 2004 (Mexico) and 2008 (Vietnam).

the provider payment, fee-for-service is still widely used in practice, which may weaken the possible financial protection effect of social health insurance due to the interaction between supply side and demand side. For the provider side, fee-for-service fundamentally gives wrong incentive to the health care provider to over-treat patients in order to pursue their own profit, such as over-prescription, taking unnecessary examinations and so on. For the demand side, when the poor are not covered by insurance, they may reject the over-treatment as it may be not affordable to them; while after they get insurance, the over-treatment turns to be cheaper to some extent so that it may cause the increase of the supply-induced demand. Empirical studies did find some evidence about this. Wagstaff et al. (2007, 2009) provided the possible explanation of non-reducing out-of-pocket health expenditure after expanding NCMS as the increase of the likelihood of people seeking medical care as well as the volume of care provided. Moreover, the statistics show a shift from lower level health facility to higher one. Most of the evaluation to Mexico's Seguro Popular shows positive financial protection to the participants. One significant difference between Seguro Popular and others -NCMS and HCFP- lies in that the former one offers free access at the point of delivery to an explicit set of health care services covering basic primary and secondary-level care as well as some high-cost tertiary-level care. Therefore, although no provider payment reform is involved in Mexico, there is no room for supply-introduced demand to the care covered in the benefit package. Furthermore, the two successful social experiments in China, which involved not only the demand side intervention but also supply side intervention (Yip & Hsiao 2009; Wagstaff & Yu 2007), may imply the importance of building up a right and balanced incentive structure in the health system reform, i.e. not only pay attention to the demand side.

One of the crucial parts of health care system reform lies in the implementation and evaluation. Mexico has explicitly planned to expand the nature and scope of monitoring and evaluation, which is financed by regular budget from the Ministry of Health. Longitudinal surveys have conducted since 2005 for the purpose of measuring the impacts of the Seguro Popular on health conditions, effective coverage, health-system responsiveness, and financial protection (Frenk et al., 2006). Ekman et al. (2008) pointed out that the capacity of policy evaluation should be further strengthened and developed in order to smooth the path to a successful health care system reform in Vietnam. The similar objection applies to China with even more serious problems. Recall that, for NCMS, the specific design of benefit package is generally decentralized to the local government (at county level). On the one hand, this decentralization is rational due to the enormous regional differences in China. On the other hand, it causes some problems, such as the absent scientificalness in benefit package design since the local governments may not have sufficient capacity to conduct policy design, nor evaluation; poor implementation due to no comparators so that there is no strong incentive for the government to improve implementation.

4.3. The eligible of social health insurance for the poor

Generally speaking, the supply side health care system reform benefits all in the sense that it aims at improving the quality and efficiency of health care service. However, targeting may be involved in the demand side reform, which is defined as identification and selection of certain groups, households or individuals and the distribution of benefits directly towards them (Mooij 1999). For the three cases we discussed in the previous sections, NCMS of China and SP of Mexico belong to the universal program in that all members of a given population are eligible to receive programme benefits: all the rural residents in China are eligible for NCMS; while all those who are not covered by any other public health schemes in Mexico, i.e. not formal employees in the public or private sectors, are eligible for SP. However, Vietnam's HCFP falls into the group of targeted program, which restricts benefits to some identified subgroups of a given population - only the identified poor and some socially protected population groups - so that it covers a much more narrowed range Hanson et al. (2008).

In most cases of social health insurance reform in the developing countries, the program is heavily subsidized by the governments. Theoretically, by focusing limited resources from public finance on those identified as having greatest needs, targeted program improves health equity in a more efficient way in that it may result in a change in individual behaviour. In our cases, NCMS of China as well as SP of Mexico is voluntary program with contribution requirement from the participants whereas HCFP of Vietnam is a compulsory program without any contribution requirement from the eligible. Apparently, HCFP should be more effective than NCMS and SP since some of the poor will not benefit from NCMS and SP due to the obligation of contribution and/or the principal of voluntary participation[11]. However, HCFP thoroughly breaks the access barrier to those who may not have been able to access the health care service due to financial problem, which leads the limited coverage range because of financial sustainability.

Where the intervention targets the poor, it raises the important issue of how to identify them. In the case of HCFP, local governments are responsible to identify the poor, many of which are identified by official poverty line. However, using poverty line as an instrument for the identification has two problems at least. First, assuming the monetary deprived are those the government wants to target, the arbitrary characteristics of the poverty line bring about the random assignment to HCFP coverage near the threshold in that the households just below or above the threshold have similar characteristics but different participation status. Second, it is debatable whether we should narrowly define the monetary poor as the poor. Nowadays, there is a heated discussion and growing literature about the multidimensional poverty measurement. Although how to measure poverty multidimensionally has not reach a common agreement, it is generally accepted that the multidimensional perspective should be taken to the poverty measurement (Sen 1985; Ravallion 1998; Gravel 2010; Alkire & Foster 2011).

[11] The poorest 20% of families are exempt from any contribution of SP.

4.4. Benefit package of social health insurance

Besides who are beneficiaries, another key component of the social health insurance is the deep of benefit package, i.e. which health care services should be included; whether the included health care services is completely free access; if not, what is the level of patient cost sharing. In terms of the range of covered health care service, there was a heated discussion in the literature (Kutzin 2000). Some literatures or policy documents suggested that the decision on which interventions to include should be based on the results from cost-effectiveness or cost-utility analysis (Eichler et al. 2004). Other literatures argued that government in the developing countries should not try a vain attempt to provide everything for everyone with its limited public finance and health resources (Baltussen 2006; Musgrove 1999). Baltussen (2006) proposed a step-wise approach to decide the targeting and prioritization of public spending. There is also some literature specially addressing the needs to concern financial protection in the benefit package design (Hammer & Berman 1995). In all, these debates do imply the trade-off between efficiency and equity in the policy design. However, in practice, these approaches do not mutually exclude. Policy maker might apply a consolidated approach and tailor it according to their specific circumstances.

SP, HCFP and NCMS differ significantly in the benefit package design. Frankly speaking, it is not obvious how they design it. SP has an explicit benefit package consisting of 249 interventions at the primary and secondary level of care and 17 others that are particularly costly, all of which are fully covered. The expansion of the package to cover more interventions with better quality is one of the essential issues in the health care system reform. Out-of-pocket health expenditure results from taking interventions excluded from the benefit package. Patient cost sharing is involved in HCFP and NCMS. The former one has a broad but undefined benefit package covering outpatient and inpatient services and even some high-tech treatments, the statistics has implied its financial unsustainability, which suggests that it might be better to provide a more focused benefit package of affordable and cost-effective interventions (Ekman 2004). However, the latter one is designed to cover mostly inpatient services rather than prevention and primary care, which gives patients the wrong incentive to over-use hospital care and under-use basic primary health care. Moreover, considering the benefit design of NCMS is decentralized to provincial and county governments under the general guidelines issued by central government, it is highly possible that the benefit package is not designed in an effective and scientific way (Yip & Hsiao 2009).

4.5. Theoretical foundation

4.5.1. The effects of insurance on the health seeking behaviour

The effects of social health insurance depend heavily on the interaction of the benefit package, the severity of disease as well as the patient's income. A representative patient's health care seeking

behaviour conditional on whether he participates health insurance or not will be discussed in this part. Note that we do not take the financing mechanism of the insurance into account. In other words, our discussion is limited to the partial equilibrium instead of general equilibrium. In fact, this partial equilibrium may make more sense due to the following reasons. There are two possible pathways to achieve general equilibrium in the context of social health insurance: (1) given the fixed price of medical care, the equilibrium could be achieved by the adjustment of the premium; (2) given the pre-set premium, the equilibrium could be achieved by the adjustment of the price of medical care. However, in most developing counties, social health insurance for the poor is heavily subsidized and supported by the government. In practice, participants pay the premium upfront and the government deals with the rest. While the surplus of the current period will be pooled into the funding of the next periods, the deficit will be coped with public finance. Recall the price is sticky in the short run, not to mention the possible price regulation or control on medical care. Hence, it is hard, at least in the short run, to achieve general equilibrium in any case.

Assume an overlapping model without borrowing and saving. The representative patient is myopic in the sense he only considers two-period utility, i.e. the current period when he falls ill (period one) and the following period (period two). The sick pay discount rate is θ $(0 < \theta < 1)$. The money spent on medical care only contributes to lower the probability of remaining the subsequent period in sickness. The patient maximizes his two-period utilities only with regard to the utility of the consumption, i.e. income net of medical care expenditure. Given the patient is rational, he spends his income on medical care M and consumptions C_1 in period one. But in period two, the income is spent only on the consumptions C_2. The average price of the medical care and consumption are p_M and p_C, respectively. In terms of the social health insurance, assuming the participation of health insurance is compulsory to all the target population and the flat premium is P in each period. That is to say, adverse selection is not taken into consideration. Denote the deductible as $\underline{R}$, copayment rate as β.

Assume that, with respect to the illness, patients can be grouped to two heads: relative rich and extremely poor. And there are two kinds of benefit packages of insurance: generous one which means positive reimbursement income effect and ungenerous one which means negative income effect. Therefore, considering the relative income of the patients and the benefit package of the insurance, our discussion will fall into the following four cases.

Case 1: extremely poor patient combined with the generous benefit package

Figure 1 shows, in the first period, the optimal allocation of consumption and medical care of the representative patient both without insurance and with insurance that pays off by reducing medical care price from p_M to βp_M if the total expenditure before reimbursement is larger than deductible. Without insurance, optimal allocation is (M^u, C_1^u), and with insurance, it is (M^i, C_1^i). An "equivalent

variation" approach is applied to decompose the total effect into income effect and price effect, in which the income effect is shown as the result of an increase in income, evaluated at the uninsured price. Thus, the price effect of insurance would be the increase in medical care demand caused by using a price reduction rather than lump sum income payments as the payoff mechanism, i.e. $(M^i - M^c)$. The income effect would be the increase in medical care caused by lump sum-payoff insurance compared with that without insurance, i.e. $(M^c - M^u)$. In this case, both the income effect and price effect are positive, which gives birth to a sharp increase of the medical care demand.

In general, Figure 1 indicates that there may be significant increase of the medical care demand in the sense that when they are uninsured, they only take a little bit treatment which is less than benefiting from reimbursement, however, after being insured, given that the benefit package is generous, they prefer to make better use of the insurance and derive the reimbursement benefit from it, i.e. medical care expenditure before reimbursement exceeds the deductible. Specifically, this could be the case that the patients are relatively poor and live on a tight budget, who only take inadequate treatment when they are uninsured. Thanks to the insurance, they have the chance to meet their basic demand. Furthermore, Figure 1 implies, although it does not show explicitly, another possible case of rising by a wide margin. That is, the patients suffering from a serious illness such as cancer forgone treatment when they are not insured, since they cannot afford such a large amount of out-of-pocket expenditure of standard medical care and it is vain to take a small quantity of medical care. After being covered by insurance, they may take a relative large quantity of medical care but pay affordable out-of-pocket expenditure. In these two cases, the extension of social health insurance is of service to give access to the patients who are in need of medical care. The clinical literature is replete with empirical studies indicating that, generally, the uninsured patients do not take standard medical care (Aday, 1993; Blendon, 1992; Cunningham, 1995). And some other studies reported that the uninsured patients tend to delay or defer seeking medical care (Ayanian, 1993; Braveman, 1993)[12].

Case 2: extremely poor patient combined with the ungenerous benefit package

It is the same as in case 1 that (M^u, C_1^u) and (M^i, C_1^i) denote the optimal allocation of income without insurance and with insurance, respectively. Due to the ungenerous benefit package, the income effect is negative. That is to say, the poor become even poorer rather than benefiting from reimbursement (Figure 2). If the patient's optimal choice is taking a small quantity of medical care without insurance, after being insured, they take less medical care with the extreme case that the patients forgone treatment. If they forgone treatment when uninsured, the extension of insurance cannot give them access to medical at all. On the contrary, paying premium leads them to have less disposable income.

[12] For the comprehensive review, please see Nyman (2003), page 21-22.

Case 3: relatively rich patient combined with the generous benefit package

If the relative rich do not take adequate medical care when there is no health insurance, the extension of generous insurance coverage gives them access to satisfy their basic demand, which is on the same principle as Figure 1. On the other hand, if the relative rich take adequate medical care even there is no health insurance (Figure 3), after being insured, there is still an increase of demand caused by the positive income effect $(M^c - M^u)$ as well as positive price effect $(M^i - M^c)$. However, it has been pointed out in the conventional health insurance demand theory that the additional health care which patients may not wish to take at full price but which they may gladly take at the lower insured price is welfare decreasing because it implies a movement along the demand curve. If marginal cost is reflected in the full price, then the additional health care is welfare decreasing[13].

Case 4: relatively rich patient combined with the ungenerous benefit package

For all the three cases discussed above, either the income effect and price effect make the difference in the same direction or only income effect plays a role in. That is to say, the total effect can be determined. Nevertheless, in the last but not least case, the income effect is negative by all means but the price effect could be positive or ineffective. Therefore, the total effect is uncertain, which depends on the extent of the ungenerosity of the benefit package. There are three possibilities as follows.

If the benefit package is ungenerous in the sense that the premium is relatively high but the copayment rate is relatively low (Figure 4), then the negative influence of income effect is dominated by the positive price effect, which means the medical care demand still increases but less than that in the generous case as the extension of insurance. However, if both the premium and copayment are sufficient high (Figure 5) and the quantity of medical care exceeds the deductible, the positive price effect may be dominated by negative income effect, which brings about a decrease, but may not so significant, of the medical care demand. The extreme case is that the benefit package is extraordinary ungenerous (Figure 6) so that the patients can not maintain their demand above the deductible, that is to say, after being insured, they are partially deprived of the access to the medical care rather than benefiting from the positive price effect.

[13] For the details about conventional theory of health insurance demand, please see Nyman (2003), chapter two.

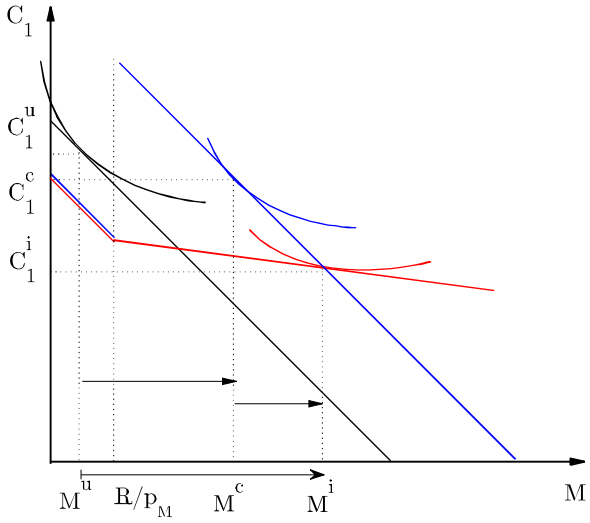


Figure 1: Extremely poor patient combined with generous benefit package

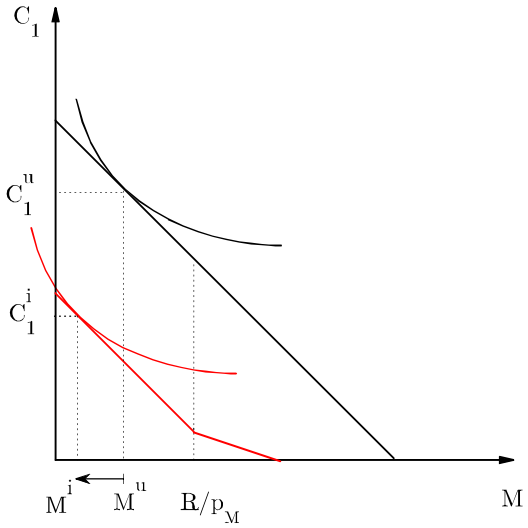


Figure 2: Extremely poor patient combined with ungenerous benefit package

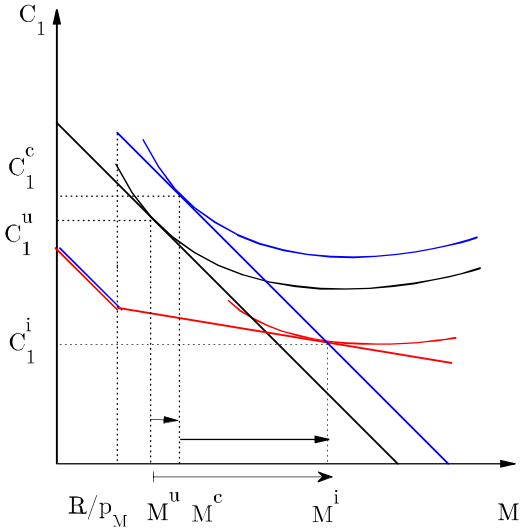


Figure 3: Relatively rich patient combined with generous benefit package

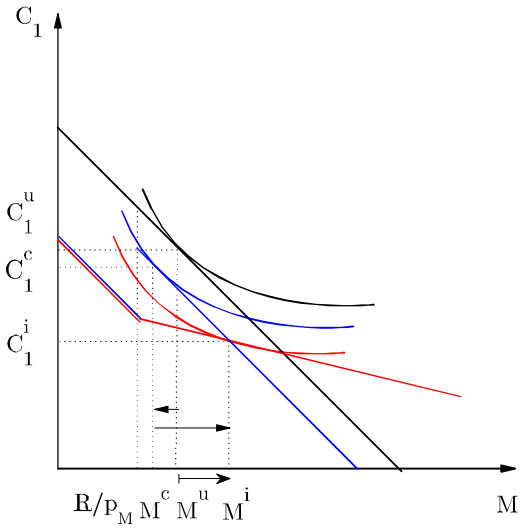


Figure 4: Relatively rich patient combined with ungenerous benefit package (a)

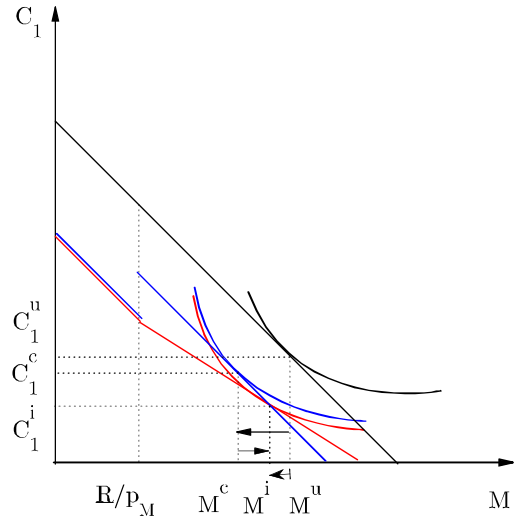


Figure 5: Relatively rich patient combined with ungenerous benefit package (b)

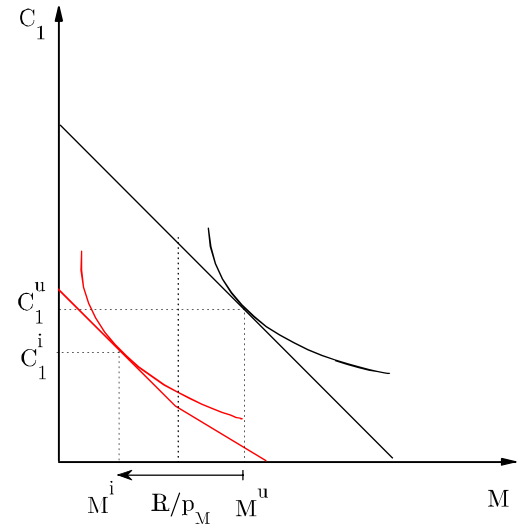


Figure 6: Relatively rich patient combined with ungenerous benefit package (c)

4.5.2. Medical poverty trap

Poverty trap is defined as the self-reinforcing mechanism which causes poverty to persist (Azariadis & Stachurski 2005). In the developing countries, many factors contributes to a poverty trap, which includes limited access to credit and capital markets, extreme environmental degradation, corrupt governance, poor education systems, disease ecology, lack of public health care, or poor infrastructure (Bonds et al. 2010). Whitehead et al. (2001) proposed a concept termed medical poverty trap to describe the negatively dynamic relationship between ill-health and poverty, which brings about four possible consequences as follows: untreated morbidity, reduced access to care, long-term impoverishment and irrational use of drugs. The following flow chart shows the pathway of medical poverty trap. It implies that for those poor patients, no matter whether they seek care or not, they may be driven into the same pathway which starts from the coping strategy due to the reduced disposable income on other consumptions.

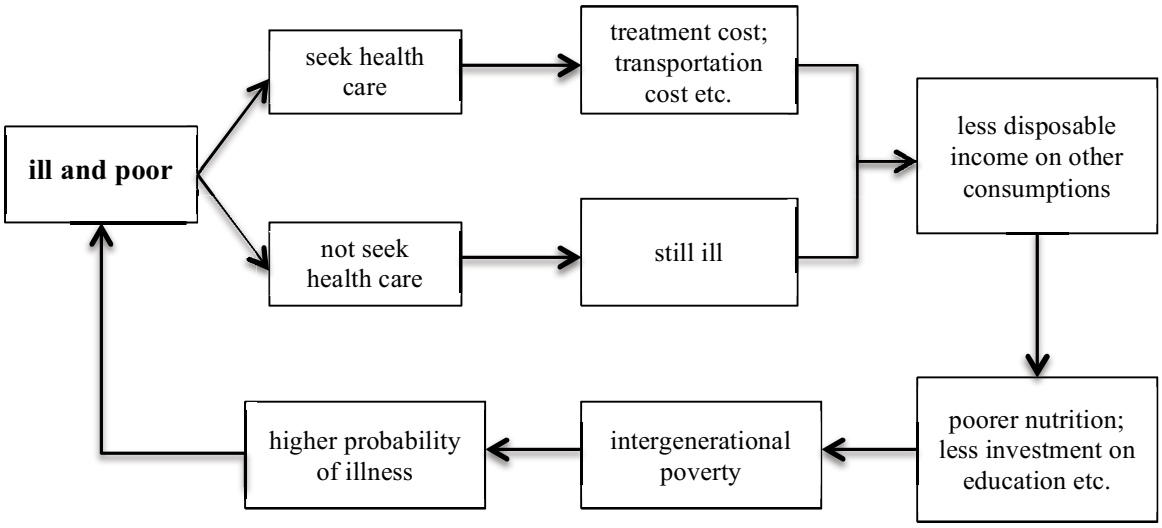


Figure 7: Pathway of medical poverty trap

This cycle may not be broken by means of introducing social health insurance. As we have discussed in the Case 2 of the previous subsection, if the benefit package is not sufficiently generous relative to the household income and health care expenditure, the poor will be pushed into an even deeper poverty trap if there is any obligated contribution. However, if the benefit package is sufficiently generous (Case 1), the ill patients will seek care without an unaffordable or over-burdened cost, they will not fall into the medical poverty trap. However, although they are lots of models describing poverty trap due to various factors, such as human capita externalities (Azariadis-Drazen 1990), child labor (Basu 1999 and Doepke and Zilibotti 2005), limited studies focus on modeling medical poverty trap, which makes this component lack of theoretical foundation.

5. Evaluation problem

Most studies about impact evaluation of social health insurance reform are empirical work. Therefore, the selection of outcome variables, the availability and quality of the data and the choice of corresponding empirical methodology given the outcome variables and data are of high relevance to the evaluation so as to the future policy design.

5.1. Outcome variables

In the literature, catastrophic health expenditure (Knaul et al., 2006; Galárraga et al., 2008; Wagstaff, 2007; Sun et al., 2009; Zhang et al., 2010) or out-of-pocket health expenditure (Axelson et al., 2009; Wu, 2010; Wagstaff et al., 2007, 2009, 2010; Babiarz et al., 2010) is often selected as the financial outcome variables. They may not be the most appropriate candidates for the policy evaluation or design due to the reasons as follows.

On the one hand, the definition of catastrophic health expenditure itself has some limitations which may mislead the policy makers. A household's health expenditure is considered to be catastrophic if the ratio between the household's out-of-pocket health expenditure and its disposable income exceeds a certain critical point. Commonly used thresholds include 30% or 40% of capacity to pay[14], or 10% of total income (Wagstaff & van Doorslaer 2003; Xu et al. 2003). One of the main limitations of catastrophic health expenditure measurement lies in the fact that it is unable to capture neither the extremely poor households who do not seek care nor those who do not seek sufficient treatment. Moreover, it does not take into account the dynamic correlation between health and income, not to mention welfare. Another limitation is that it only focuses on the potentially financial risk due to out-of-pocket health expenditure in a short run without considering the heterogeneous copying ability of the households. Specifically, if introducing social health insurance improves the health utilization of the poor but some of their out-of-pocket health expenditure share may exceed some arbitrary threshold, they will be accounted as falling into the group of catastrophic health expenditure, which may lead the negative evaluation results with regards to financial protection function of social health insurance. However, they could financially benefit from introducing social health insurance in the long run due to the positive contribution of health to income. For the some out-of-pocket health expenditure share, the relative rich could be much better-off comparing with the poor since the rich have more feasible coping strategy.

On the other hand, the ultimate goal of the health care system reform or promoting social health insurance is not to neither lower the financial risk of out-of-pocket health expenditure nor increase

[14] "Capacity to pay" is defined as effective income net of basic subsistence needs have been met. In practice, effective income is taken to be the total consumption of the household since in many countries it is a more accurate reflection of purchasing power than income reported in household surveys; basic subsistence needs is taken to be the total consumption spending on food.

health care utilization. Ruger (2010) addressed that "clearly wealth is not the good we are seeking"; "human flourishing is the end of all social activity". That is to say, it could be shortsighted to evaluate health policy from single separated outcome variables. Instead, it could be better if we jointly consider and evaluate the multiple impact of social health insurance within a well-being framework which embodies not only the income dimension but also health-related dimensions, since, as we have discussed before, health and income dynamically influence each other but jointly contribute to the overall well-being. Note that here we accept the notion that well-being should be defined from a multidimensional perspective, i.e. well-being goes beyond the pure monetary item or a specific type of consumption good (Haughton & Khandker 2009; Stiglitz et al. 2009; Sen 1985).

5.2. Data and methodology

Chapman & Boothroyd (1988) argued that the quality of data employed in the analysis is frequently overlooked. Especially under the setting of developing countries, data quality may be negatively affected due to external researchers' limited experience or knowledge of local conditions. Survey data are mainly employed in the health care system evaluation. Survey error may arise from frame deficiencies, sampling process, interviewing and interviewers, respondents, missing data, coding and so on (Biemer 2010). However, without ensuring the quality of the data, the findings of analysis can be biased or misleading regardless how sophisticated methodology is applied, which may consequently result in the random decisions of policy makers. While most of the evaluation analysis uses data from general household survey except for the evaluation on some social experiment, Mexico is leading the health system reform in the developing countries also in that it has been conducting special-designed longitudinal survey from 2005, which could provide well-tailored information to evaluate the design and implementation of SP so as to benefit the further policy design.

Most of the policy evaluation studies are conducted by non-experimental data. Consequently, a wide variety of estimation methodologies have been proposed and discussed in the literature to overcome the selection bias problem in non-experimental evaluation. Matching, difference-in-difference and instrumental variables are most widely used. However, there is an ongoing debate in the literature about whether policy or interventions can be reliably evaluated without conducting a randomized experiment (LaLonde 1986; Smith & Todd 2005). The essential advantage of randomization experiment lies in that we can generate a control group that has the same distributions of both observed and unobserved characteristics as the treatment group so that the selection bias problem can be naturally solved. Nevertheless, compared with non-experimental evaluation, randomization experimentation generally faces the problem of high cost. In addition, its validity and reliability is determined by the commitment of treatment groups and control groups.

In terms of the evaluation on health care system reform in the developing countries, the majority of the research employs non-experimental data except for some small scope social experiment. There is a

growing evidence shows that impact estimates are often highly sensitive to the estimation methodology (Smith & Todd 2005). Difference-in-differences, a quasi-experimental technique that measures the effect of a treatment at a given period, has been extensively applied in health policy evaluation. Difference-in-difference estimator represents the difference in an outcome before and after the treatment between the treatment and control groups, which implies the requirement of panel data or repeated cross-section data as well as the specification of treatment and control groups. The difference-in-difference estimator combines selection on time-invariant unobservables (by individual-effect variable) with selection on observables (by covariates). The most strict assumption[15] of difference-in-difference is the parallel trend assumption, i.e. all the time-varying effects are common to both the treatment and control groups (Wang et al. 2009). The parallel trend assumption can be investigated if multiple period data are available (Card & Krueger 2000). In practice, many studies enhance the comparability of treatment and control groups by combining the difference-in-difference approach with matching. However, matching provides a general approach to overcome the observed differences between treatment and control with the assumption that there is no unobserved difference, which seems implausible.

6. Concluding remarks

In this chapter, we discuss health care system reform in the developing countries with the special attention to their impact on the poor. SP of Mexico and HCFP of Vietnam are chosen as representatives to compare with NCMS of China to explain why most studies do not report the positive impact in the case of China. While Mexico could be regarded as one of the most successful cases in the health care system reform in developing countries, Vietnam has more common ground in the demographic factors with China.

Our discussion focuses on two aspects: policy design and evaluation methodology. In terms of the policy design, background differences, identification of beneficiary, design of benefit package and theoretical foundation are discussed separately. However, it is possible to comment on the single policy design to some extent, whereas it is actually difficult or implausible to judge the policy package as a whole since countries have different starting points and governments always face various constraints and tradeoff. Explicitly speaking, although Mexico has got great achievement in the reform of health care system, we should be cautious to apply their experience with regard to the detailed policy design. In terms of the evaluation-related issues, the selection of outcome variables, the availability and quality of the data and the choice of corresponding empirical methodology given the outcome variables and data are of high relevance. It is worth noting that when non-experimental data

[15] When difference-in-difference is applied to repeated cross-section data, a further assumption that the composition of the treatment and control groups remains stable over time is required.

are employed, we should be cautious to choose the estimation methodology and interpret the results. Before applying a methodology, it is of significant importance to check whether the assumptions are true or not. It is not appropriate to just apply the popular methodology in the literature without checking the suitability and validity under the specific circumstance. Otherwise, the findings will be biased and misleading so as to have negative influence to the future policy design.

Most of the ongoing health care system reforms focus on the demand side intervention, which aims at extending social health insurance to the non-insured and/or providing more generous benefit package to the poor so as to improve equal access to the basic health care as well as provide financial protection. Nevertheless, both economic theory and empirical evidence have suggested that health insurance itself may have limited influence on reducing patients' financial risk, which are caused by the interaction of patients' incentive to increase health care demand as well as providers' incentive to increase the level of demand inducement. Currently, most of the supply side interventions are still at the stage of small-scale social experiment. Therefore, it is of importance to give more attention to the supply side intervention such as provider payment systems, which plays a crucial role in the success or failure of health care system reform.

Reference

Alkire, S. & Foster, J., (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95(7-8)*,476-487.

Axelson, H. et al., (2009). Health financing for the poor produces promising short-term effects on utilization and out-of-pocket expenditure: evidence from Vietnam. *International journal for equity in health, 8(20)*.

Babiarz, K.S. et al., (2010). New evidence on the impact of China's New Rural Cooperative Medical Scheme and its implications for rural primary healthcare: multivariate difference-in-difference analysis. *Bmj, 341(c5617)*.

Baltussen, R., (2006). Priority setting of public spending in developing countries : do not try to do everything for everybody. *Health Policy, 78,*149-156.

Biemer, P.P., (2010). Total survey error: design, implementation, and evaluation. *Public Opinion Quarterly, 74(5)*, 817-848.

Card, D. & Krueger, A.B., (2000). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania: Reply. *American Economic Review, 90(5)*, 1397-1420.

Chapman, D.W. & Boothroyd, R.A., (1988). Threats to data quality in developing country settings. *Comparative Education Review, 32(4)*, 416-429.

Ekman, B., (2004). Community-based health insurance in low-income countries: a systematic review of the evidence. *Health Policy and Planning, 19(5),* 249-270.

Ekman, Björn et al., (2008). Health insurance reform in Vietnam: a review of recent developments and future challenges. *Health policy and planning, 23(4)*, 252-263.

Frenk, J. et al., (2006). Comprehensive reform to improve health system performance in Mexico. *The Lancet, 368(9546),* 1524-1534.

Frenk, J. et al., (2003). Evidence-based health policy: three generations of reform in Mexico. *The Lancet, 362,* 1667–1671.

Galárraga, O. et al., (2008). The impact of universal health insurance on catastrophic and out-of-pocket health expenditures in Mexico: a model with endogeneous treatment. *HEDG Working Paper, 12*.

Galárraga, O. et al., (2010). Health insurance for the poor: impact on catastrophic and out-of-pocket health expenditures in Mexico. *The European journal of health economics, 11(5)*, 437-447.

Gravel, N., (2010). Nanak Kakwani and Jacques Silber (eds): Quantitative approaches to multidimensionnal poverty measurement. *Social Choice and Welfare, 35(2)*, 337-343.

Hanson, K., Worrall, E. & Wiseman, V., (2008). Targeting services towards the poor: A review of targeting mechanisms and their effectiveness. In *Health, economic development and household poverty: from understanding to action*. London: Routledge.

Haughton, J. & Khandker, S.R., (2009). *Handbook on Poverty and Inequality*, The World Bank.

Jones, A.M., (2007). Panel data methods and applications to health economics. *HEDG Working Paper 07/18*, University of York.

King, G. et al., (2009). Public policy for the poor? a randomised assessment of the Mexican universal health insurance programme. *The Lancet, 373(9673),* 1447-1454.

Knaul, F.M. et al., (2006). Evidence is good for your health system: policy reform to remedy catastrophic and impoverishing health spending in Mexico. *The Lancet, 368(9549),* 1828-1841.

LaLonde, R., (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review, 76,* 604-620.

Lei, X. & Lin, W., (2009). The new cooperative medical scheme in rural China: does more coverage mean more service and better health? *Health Economics, 18,* S25-S46.

Long, G.T., (2007). Social health insurance in Vietnam: current issues and policy recommendations. I*LO Series: Social Security Extension initatives in East Asia.*

Mooij, J., (1999). Real targeting: the case of food distribution in India. *Food Policy, 24,* 49-69.

Musgrove, P., (1999). Public spending on health care: how are different criteria related? *Health Policy, 47,* 207-223.

Ravallion, M., (1998). Poverty lines in theory and practice: Living standards measurement study. *The World bank Working Paper No.133*.

Ruger, J.P., (2010). *Health and Social Justice,* New York: Oxford University Press.

Sen, A., (1985). Commodities and Capabilities, North-Holland.

Smith, J.A. & Todd, P.E., (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics, 125(1-2),* 305-353.

Stiglitz, J., Sen, A. & Fitoussi, J., (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

Sun, X. et al., (2009). Catastrophic medical payment and financial protection in rural China: evidence from the New Cooperative Medical Scheme in Shandong Province. *Health Economics, 18,* 103-119.

WHO, (2000). Health systems: improving performance. *World Health Report.*

WHO, (2006). Working together for health. *World Health Report.*

Wagstaff, A., (2010). Estimating health insurance impacts under unobserved heterogeneity: the case of Vietnam's Health Care Fund for the Poor. *Health Economics, 19,* 189-208.

Wagstaff, A., (2007). Health insurance for the poor : initial impacts of Vietnam ' s Health Care Fund for the poor. *World Bank Policy Research Working Paper, 4134.*

Wagstaff, A. & van Doorslaer, E., (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Economics, 12(11),* 921-934.

Wagstaff, A. & Yu, S., (2007). Do health sector reforms have their intended impacts? The World Bank's Health VIII project in Gansu province, China. *Journal of Health Economics, 26(3)*, 505-535.

Wagstaff, A. et al., (2009). Extending health insurance to the rural population: an impact evaluation of China's new cooperative medical scheme. *Journal of Health Economics, 28(1)*, 1-19.

Whitehead, M., Dahlgren, G. & Evans, T., (2001). Equity and health sector reforms: can low-income countries escape the medical poverty trap? *The Lancet, 358(9284),* 833-836.

Wu, B., (2010). The effects of the health insurance availability on the demand-side: an impact evaluation of China's New Cooperative Medical Scheme. *NIFS Working Paper No. 000017*, Tsinghua University.

Xu, K. et al., (2003). Household catastrophic health expenditure: a multicountry analysis. *The Lancet, 362(9378)*, 111-117.

Yi, Hongmei et al., (2009). Health insurance and catastrophic illness : a report on the New Cooperative Medical system in rural China. *Health Economics, 18,* S119-S127.

Yip, W. & Hsiao, W., (2011). China's health care reform : a tentative assessment. *China Economic Review, 20(4)*, 613-619.

Yip, W. & Hsiao, W.C., (2009). Non-evidence-based policy: how effective is China's new cooperative medical scheme in reducing medical impoverishment? *Social science & medicine, 68(2)*, 201-209.

Zhang, L. et al., (2010). How effectively can the New Cooperative Medical Scheme reduce catastrophic health expenditure for the poor and non-poor in rural China? *Tropical Medicine & International Health, 15(4)*, 468-475.

Catastrophic Health Expenditure and Well-being In a Multidimensional Poverty Framework

Abstract

This chapter proposes to combine catastrophic health expenditure literature with multidimensional poverty literature to analyse the impact of out-of-pocket health expenditure on overall well-being. An alternative definition of catastrophic health expenditure is introduced. Moreover, out-of-pocket health expenditure is measured not only from a positive perspective but also from a normative perspective to take the underutilization and overutilization of health care into account. The arguments and methods are empirically illustrated with data from 2009 Ningxia (China) Household Health Survey data. Our study suggests that, in the rural area of developing countries, especially where health care system is in its infancy, households may be driven into poverty by health-related deprivation more than monetary deprivation. Therefore, policy-makers should evaluate and design welfare policy from a broader perspective other than only focusing on addressing the monetary poverty.

***Keywords*:** multidimensional poverty; catastrophic health expenditure; affordability

1. Introduction

It has been argued that in the absence of well-developed health system, households' welfare may be severely disrupted by spending a large fraction of the their budget on health care (Berki, 1986; van Doorslaer et al., 2007). In the empirical studies, the conventional way to capture how households' well-being is affected by out-of-pocket health expenditure is to assess the incidence and degree of catastrophic health expenditure. A household's health expenditure is considered to be catastrophic if the ratio between the household's out-of-pocket health expenditure and its disposable income exceeds a certain critical point. Commonly used thresholds include 30% or 40% of capacity to pay[16], or 10% of total income (Wagstaff & van Doorslaer, 2003; Xu et al., 2003). One of the main limitations of catastrophic health expenditure measurement lies in the fact that it ignores the extremely monetary poor households who forgo treatment. These households probably suffer an even greater well-being loss than those who are identified as incurring catastrophic health expenditure. Another limitation is that it only focuses on the potentially financial risk due to out-of-pocket health expenditure in a short run without considering the possible coping strategy of the households as well as the return of health care.

In this chapter, we study the impact of out-of-pocket health expenditure on overall well-being rather than only on potentially financial risk, which raises a key question about how to define well-being. In general, there are three perspectives of defining well-being and the corresponding poverty line (Haughton & Khandker, 2009). First, in a most conventional view, well-being is seen largely in monetary terms such as income or consumption. In this case, being deprived in well-being means income or consumption below some given monetary poverty line. Second, well-being focuses on a specific type of consumption good, such as nutrition or education, which goes beyond the monetary view. In this case, deprivation can be determined by convention since common-agreed standards are usually available. Another broader view – known as Sen's capability approach - argues that well-being stems from capability to function well in society, according to which, poverty arises when people are deficient in key capabilities. Sen's capability approach is rooted in Aristotelian view, which argues, "clearly wealth is not the good we are seeking"; "human flourishing is the end of all social activity" (Ruger, 2010). In the practice of measurement, well-being consists of, at least in principle, the following dimensions: (1) material living standards (income, consumption and wealth); (2) health; (3) education; (4) personal activities including work; (5) political voice and governance; (6) social connections and relationships; (7) environment (present and future conditions); (8) insecurity, of an economic as well as a physical nature (Stiglitz et al., 2009). Although not all the researchers and policy makers agree with Sen's emphasis on the concepts of 'capability' and 'functionings', they have

[16] "Capacity to pay" is defined as effective income net of basic subsistence needs have been met. In practice, effective income is taken to be the total consumption of the household since in many countries it is a more accurate reflection of purchasing power than income reported in household surveys; basic subsistence needs is taken to be the total consumption spending on food.

generally accepted the notion that the multidimensional perspective should be taken to the poverty measurement. In our analysis, well-being will be measured as the combination of five dimensions: disposable income, living standards, education, health status and health expenditure.

We combine catastrophic health expenditure literature with multidimensional poverty measurement proposed by Alkire & Foster (2011a) to analyse the impact of out-of-pocket health expenditure on overall well-being. This combination brings about at least two benefits.

On the one hand, it overcomes the respective limitations or possible misunderstandings presented when each approach is applied alone. It reveals the households who are too monetary poor to utilize health care and hides those who currently spend a lot on health care but it is 'affordable' for them by the indicators of our multidimensional framework, such as income, living standards and health status. In terms of overcoming the limitations of multidimensional poverty framework, given the empirical evidence about catastrophic health expenditure, if we measure multidimensional poverty without taking out-of-pocket health expenditure into account, the overall poverty may be significantly underestimated.

On the other hand, it sheds lights on the appropriate distribution of public finance. Many developing countries are implementing health care system reform with the support from public finance. However, there is little guidance about how much should be spent on health as opposed to other sectors in economic theory. Thanks to Alkire and Foster (2011), we can measure relative contribution of different dimensions to the overall poverty. Hence, policy-makers can look health expenditure issue from a broadened framework, which gives suggestions to them about whether they should concentrate on improving health care system or there is something else on a higher priority to promote people's well-being with limited resources.

In our study, health expenditure is considered as one of the dimensions together with income, living standards, education and health status. After the poor are identified, the property of dimensional decomposition can be applied to obtain the percentage contribution of each dimension to overall poverty, the comparison among which indicates the priority of government policy. Furthermore, in terms of policy evaluation, it is of some relevance to compare the welfare gain due to the policy intervention with other welfare gains obtained with no intervention.

Besides the proposed combinational method, there are some other contributions of this study. First, literatures on measuring affordability of housing and utility (Kessides et al., 2009; Thalmann, 2003) are introduced to broaden the concept of catastrophic health (Wagstaff & van Doorslaer, 2003, Xu et al., 2003). Beside the standard catastrophic health expenditure definition, i.e. budget share approach from the view of affordability literature, residual income approach is applied in our analysis. Second, we propose to measure out-of-pocket health expenditure from positive and normative perspectives,

respectively, to take the underutilization and overutilization of health care into account. It is aware that the normative expenditure is reckoned roughly due to some data unavailability. However, it is still worthwhile doing it since the potential problems that should be taken into account in the further research and policy design could be exposed. Comparisons are made between different definitions of catastrophic health expenditures as well as perspectives. Furthermore, how data resources possibly bias the results is discussed, in which aggregated health expenditure data from consumption module and disaggregated data from health module are involved.

Our method is applied to a data set collected from the rural area of Ningxia province (China) to analyze the impact of out-of-pocket health expenditure to overall well-being. As is in other countries, catastrophic health expenditure is conventionally used in the literature related to China to indicate the impact of out-of-pocket health expenditure on households' welfare (van Doorslaer et al., 2007; Sun et al., 2009; Wagstaff & Lindelow, 2008; Zhang et al., 2010). Furthermore, there are very few published literatures applying multidimensional poverty framework to measure the poverty in China (Labar & Bresson, 2011). Therefore, this paper also contributes to apply multidimensional poverty framework to China data.

The paper is organised as follows. Section 2 provides a survey on main poverty measurement and gives the rationale of choosing Alkire and Foster (2011) as our fundamental framework. Section 3 presents the methodology that is applied in our study. Section 4 introduces data resources, and sampling design. Section 5 describes the dimensions, indicators and cutoffs involved in this study, in which the indicator related to measuring catastrophic health expenditure is discussed in detail. Section 6 briefly introduces the weighting structure. Section 7 reports the findings related to identification procedure and aggregation procedure.. Also, subgroup decomposition and dimensional decomposition are conducted under the benchmark settings. Section 9 checks the robustness of the results. Section 10 concludes. Some relevant but not key tables, graphs and proofs are attached in the appendix.

2. Literature review

In the empirical studies, poverty measurements can be classified into two categories: unidimensional measurements and multidimensional measurements.

2.1. Unidimensional measurements

Unidimensional measurements usually select income or total expenditure[17] as the indicator. Given the selected indicator, the problem exists in the sense that it is hard to determine a clear threshold to distinguish between the poor and non-poor. Silber (2007) referred this cutoff problem as "fuzzy aspect" of poverty. In the literature, there are two different sorts of approaches to determine poverty cutoffs: objective and subjective approaches.

Objective approaches assume that there is a basic need of each household to lead a healthy and normal life. Within this sort, a distinction is made between absolute and relative poverty lines. The former has a fixed value over time and space while the latter is influenced by the reference level of expenditure or income (Ravallion, 1998). To interpret from the definitions of poverty, the former means having less than an objective defined, absolute minimum while the latter implies having less than others in society (Hagenaars & de Vos, 1988). Differentiated by some selected characteristics such as family size, the sex of the family head, the number of family members who were children, Orshansky (1965) defined an absolute poverty matrix composed of 124 elements to identify the poor. However, Townsend (1979) criticized the definitions in terms of absolute level of minimum needs as inappropriate and misleading, on the basis that needs are conditioned by the society where people live. Hence, he proposed a poverty measurement labelled as relative poverty line in order to capture the social influence.

The objective approaches are not truly objective given the fact that the value judgments affect measurements. In other words, who are the judgment-makers matters. In most cases, they are those who are carrying out the research. Possibly, the basic needs of households in their mind are not in conformity with what in the households' mind. Subjective approaches have been put forward to respond to this issue[18]. In contrast with having the researchers defining the basic needs, bearing the risk to misinterpret what the basic needs of households are, this approach designs *ad hoc* minimum income questions in the household survey. That is to say, the respondents of the survey are in charge of defining their own basic needs. For instance, "What is the smallest amount of money a family of four (husband, wife, and two children) needs each week to get along in this community?" (Kilpatrick, 1973). If their actual income is less than the amount that they consider being 'just sufficient', they are identified as the poor. In this sense, poverty is defined as the feeling that you do not have enough to get along (Hagenaars & de Vos, 1988). However, this would give birth to inconsistency in the resulting poverty measures, in the sense that household with the same income may give different

[17] There were some debate about which was more appropriate. Some studies rely on the total expenditure rather than income since there is a lack of reliability of self-reported income in surveys and goods as well as services that are provided on a subsidized basis are not taken into consideration.

[18] Definition of this approach was developed by Goedhart et al. (1977). Applications and extensions can be found in Van Praag et al. (1980, 1982), Danziger et al. (1984) etc.

answer to the minimum income question so that they will not be considered equally poor (Ravallion 1998)[19].

Sen (1976) pointed out two principal problems in the poverty measurement, which could be summarized as identification and aggregation. The discussion above on how to determine poverty line reflects the main identification issue in the unidimensional poverty analysis. Once the poor have been identified, an aggregation method needs to be established so that the poverty status of each household can be summed up to a society level, which is of more significance as well as importance to the policy-makers. The aggregation issue of the unidimensional poverty measurement will be elaborated in the following multidimensional part given that it will cover all the problems that may come along here.

2.2. Multidimensional measurements

Researchers (Nussbaum, 2001; Sen, 1985) advocated that the traditional unidimensional indicator - often income or total expenditure - cannot capture the definition of poverty well since it keeps blind to the information about life quality, which should be considered in the poverty measurement.

2.2.1. The challenges in multidimensional poverty analysis

Compared with the unidimensional measurements, at least three additional difficulties emerging in the multidimensional measurements are: the choice of dimensions, the cutoff across the dimensions, weighting structure. Clark & Qizilbash (2005) labelled the first two as "horizontal vagueness" and "vertical vagueness", respectively.

With respect to the choice of dimensions, the following questions must be taken into account (Silber, 2007): which dimensions are relevant and what are their interactions; for a given dimension, which indicators should be selected and what are their interactions. Alkire (2008) summed up five possible means of selection as follows: 1) rely on deliberative participatory process; 2) follow a list that has achieved certain common agreement at the international level; 3) make assumptions from a normative perspective; 4) consider data availability or authoritative convention; 5) draw from empirical evidence. These methods can be applied either alone or in combination.

In addition to this horizontal vagueness problem, vertical vagueness may arise in the context of multiple dimensions, i.e. differentiating the poor from the non-poor becomes more questionable. Who should be labelled as poor? For instance, Atkinson (2003) first discussed an intuitive counting approach and applied the terms "intersection" and "union" to the two extremes -poor in all the

[19] Past empirical studies suggest that the answer to the minimum income question conditional on actual income trends to be an increasing function of actual income. Also, family size and some demographic components may have influence on the subjective poverty line (Ravallion, 1998).

dimensions and poor in at least one dimension, respectively- while Alkire and Foster (2011) proposed to use an intermediate cutoff.

Moreover, one of the crucial steps in constructing the multidimensional measurements is the choice of the relative weights for each dimension. Munda & Nardo (2005) pointed out that when the linear aggregation was applied, weights could be seen as representing the importance of the dimension to the overall well-being. Decancq & Lugo (2010) critically surveyed the eight different approaches to set weights, which were classified into three categories: data-driven weights, hybrid weights and normative weights. After comparing pros and cons of each weighting scheme, they concluded that, given that there is no widely agreed theoretical framework on how to set weights, researchers have no choice but to rely on their common sense. Therefore, we should be cautious in interpreting the results and, sensitivity analysis is necessary in any case.

2.2.2. The main approaches

The main approaches to multidimensional poverty measures are latent variables, fuzzy set, information theory, distance function and axiomatic derivation.

Latent variables approach assumes that the characteristics of some unobservable variables such as poverty can be reflected by a set of observable variables called indicators. By means of the traditional multivariate regression, it estimates the poverty using the information provided by the selected indicators. Factor analysis, multiple indicators and multiple causes model and structural equation model could be three representatives of this approach[20]. Factor analysis (Maasoumi & Nickelsburg, 1988; Schokkaert & van Ootegem, 1990), the simplest latent variable model, relies on the assumption that the observed variables are linear combinations of certain common latent variables, known as "factors". In terms of Sen's capability approach, it explains the functionings by means of capabilities represented by the latent factors. However, it does not explain what causes these latent variables to change. The multiple indicators and multiple causes model (Kuklys, 2005; Di Tommaso, 2006) bridges this gap by integrating the exogenous variables that influence the latent variables into the model. Specifically, there are two linear combinations in this model: the observed variables are postulated to be linear functions of some latent variables and the latent variables are postulated to be linear functions of some causes. Nevertheless, this model is still quite limited since it only builds a one-way causal link. Structural equation model (Batana, 2010; Krishnakumar & Ballon, 2008; Krishnakumar, 2007; Wagle, 2005; Wagle, 2007) goes further in the sense that it gives a more complete picture of the mechanism. Two main components of this model are: the structural component showing potential causal dependencies between latent endogenous and exogenous

[20] In addition to these models mentioned above, principle component analysis is also widely used in the measurement of well-being. It is more a data reduction and dimension aggregation technique than a method based on the idea of latent variables. For a detailed discussion about latent variables, see Kakwani and Silber (Eds.): Quantitative approaches to multidimensional poverty measurement (2008), chapter 3-7, Palgrave Macmillan.

variables, and the measurement component showing the relations between latent variables and their indicators. Krishnakumar & Nagar (2007) made the motivation of these approaches[21] explicitly and checked their suitability in the economic and social domain.

In the multidimensional poverty measurement, it is even more difficult to determine a clear threshold to make a difference between those who are poor and those who are not. The fuzzy set approach proposed by Zadeh (1965) is based on the idea that certain classes of objects may not be defined by very precise criteria of membership. He characterized a fuzzy set as "a class with a continuum of grades of membership." Cerioli & Zani (1990) first applied the concept of fuzzy set to the measurement of poverty and proposed so-called total fuzzy approach. Cheli & Lemmi (1995) developed Cerioli and Zani's approach in the fact that it - totally fuzzy and relative approach- does not make use of any critical threshold values so that it is less arbitrary. Also, it takes into account the distribution of each indicator in the society, i.e. define poverty by a relative approach. The totally fuzzy and relative approach consists of three steps: 1) define the deprivation level of individual i with respect to dimension j, which equals to the proportion of individuals who are not more deprived than he is; 2) compute the overall level of deprivation of each individual; 3) aggregate poverty information from individual level to population level. However, Vero & Werquin (1997) noted that some of the indicators/dimensions may be highly correlated. Thus, they defined a new membership function to overcome this problem[22].

Information theory approach is based on the idea that data can be transmitted without significant losses or errors by means of the relative entropy measure (Shannon, 1948). That means, in the context of multidimensional measurement, it is feasible to derive an individual aggregate function whose distribution is the least divergent from the distribution of the dimensions. To measure inequality, Maasoumi (1986) applied this approach both in the aggregation across the dimensions to obtain the index for each individual and in the aggregation across individuals to obtain the overall inequality. On the basis of Maasoumi (1986), Lugo & Maasoumi (2008) introduced this approach to poverty measurement. They made a distinction between two alternative families of measures depending on the underlying definition of the poverty line, which were termed aggregate poverty line measure and component poverty line measure, respectively. Just as their names imply, the former one aggregates the information across the dimensions and then compares it with the aggregate poverty line for each individual while the latter one firstly computes the shortfall of each dimension for each individual and then aggregates the shortfalls across the dimensions for each individual. The following step, the same for both alternatives, is to define a set of the poor and aggregate the poverty across the individuals. In

[21] Principle component analysis is included in the comparison.
[22] For the detailed discussion about fuzzy set approach, see Lemmi and Betti (Eds.): Fuzzy Set Approach to Multidimensional Poverty Measurement (2006), Springer.

addition, they differentiated between the strong and weak poverty focuses, which are distinguished by whether some substitution between the dimensions above and below the poverty line is allowed[23].

All the approaches presented above are based on the inductive reasoning, in which way the general conclusions are drew from the observed patterns under a set of hypothesis put forward beforehand. Therefore, it is not a valid method of proof to the extent that uncertainty remains regardless of how many observations are made. On the contrary, deductive reasoning is the step-by-step process of gathering and organizing the known information to reach conclusions on the basis of some clear and self-evident principles. So that it is often promoted as able to tell us what we are really getting.

In the multidimensional poverty literature, (Chakravarty et al., 1998; Tsui, 2002) were the first who made attempts to derive poverty indices by deductive reasoning, which is referred to as axiomatic approach. The basic idea behind is as follows. Multidimensional poverty indices aggregate the shortfalls of all the individual observations with respect to some pre-specific minimum levels of needs, which size up the situation of the identified poor. The desirable properties proposed by Chakravarty et al. (1998) include the following: symmetry, focus, monotonicity, principle of population, continuity, non-poverty growth, non-decrease in subsistence levels of basic needs, scale invariance, normalization, subgroup decomposability, factor decomposability, transfer axiom as well as non-decreasing poverty under correlation increasing arrangement. Afterwards, Bourguignon & Chakravarty (2003) argued that, if the various dimensions cannot be aggregated into an overall variable in a natural way, dimension-specific poverty line should be used. Then they discussed the general forms of identification functions rather than put forward an axiomatic structure for it. In other words, overall methodology that could be of help to build up new identification techniques in practice not is well covered in this literature.

The latest development of axiomatic approach for constructing multidimensional poverty measurement is promoted by Alkire and Foster (2011). In this literature, they built up a general two-step framework, guided by which, poverty indices satisfying a range of desirable properties can be produced. The attractiveness of this framework to our study lies in the following aspects. First, it is well suited to work not only with continuous and cardinal data but also with discrete and ordinal data; second, it is appropriate when dimensions are fundamentally incommensurate achievement and cannot perfectly compensate each other; third, it is widely flexible and adaptable in the sense that it does direct the way to construct indices with desirable properties but leaves the specific components, such as dimensions, weights and cutoffs, to be chosen in accordance with the purpose of measurement and data resources. A full picture of this framework will be drawn in the following section.

[23] Weak focus makes the poverty index independent of the attribute levels of non-poor individuals only. In contrast, Strong focus leads us to ignore not only individuals above the poverty threshold in all relevant attributes, but also attributes above the poverty threshold for individuals who do not achieve the threshold in other attributes. For the detailed discussion about information theory approach, see Lugo and Maasoumi (2008).

3. Methodology

Sen (1976) concisely summarized two problems that must be faced in the poverty measurement: (1) identification problem, i.e. how to choose the criterion of poverty and then distinguish those who fall into that criterion and those who do not;, and (2) aggregation problem, i.e. how to construct an index of poverty using the available information on the poor. The multidimensional poverty framework proposed by Alkire and Foster (2011) guides the way to overcome these two problems[24].

This framework consists of five key components: dimensions, indicators, deprivation cutoffs, poverty cutoff and weighting structure. Well-being is measured by selected dimensions, which are income, living standards, education, health status and health expenditure. For each dimension, several indicators are selected to represent this dimension. For instance, the access to flushing toilet and tap water could serve as the indicators of living standards dimension. The number of the indicators using in a given dimension is subject to the data availability and research purpose. Each indicator has a corresponding deprivation cutoff, which is a threshold determining whether the household is deprived with respect to a given indicator. In terms of the weighting structure, we need to consider the weights within each dimension and across the dimensions. Given a selected weighting structure, all the deprivation status of each single household can be aggregated to the overall deprivation of this household. Poverty cutoff is the threshold determining whether the household is poor. To be specific, a given household is identified as the poor if and only if its overall deprivation exceeds the poverty cutoff.

The unit of the analysis is household. Let n represent the number of households. Denote d as the number of dimensions, while d_j as the number of indicators in dimension j. y_{jk}^i denotes the achievement of household i on the k-th indicator belonging to dimension j and z_{jk} denotes the corresponding deprivation cutoff. In terms of the weighting structure, recall that there is within dimension weights as well as across dimension weights. Let w_{jk} be the within-dimension weight of the k-th indicator belonging to dimension j and w_j be the across-dimension weight of dimension j. Therefore, the overall weight of k-th indicator belonging to dimension j, denoted as $w_{(jk)}$, is the product between w_{jk} and w_j. For simplicity, we assume: (1) for any given dimension, within-dimension weights sum up to 1, i.e. $\sum_{k=1}^{d_j} w_{jk} = 1, \forall j = 1,2,\dots d$; (2) across-dimension weights sum up to the number of dimensions, i.e. $\sum_{j=1}^{d} w_j = d$. The denotations of deprivation cutoff and weights

[24] We basically follow Alkire and Foster (2011) but there are some differences. Indicators (so as to within-dimension weights) were not involved in their introduction of methodology, nor the illustrative examples. So they had three-step identification procedure: (1) apply dimensional-specific deprivation cutoffs to each dimension; (2) aggregate dimensional deprivation to overall deprivation; and (3) employ the poverty cutoff to identify the poor. However, indicators are employed to represent each dimension in our study. Therefore, our identification procedure is extended to four steps which will be described in detail. Note that indicators did play a role in Alkire and Santos (2010) and Foster (2007); however, there were an inconsistency between the methodology and data analysis in their papers in the sense that they discussed the methodology at the level of dimensions but conducted analysis at the level of indicators.

imply that we assume deprivation cutoffs and weights of each dimension are the same for every household. However, in practice, the individual household's characteristics can be taken into account through appropriate equivalence scales. We conduct a four-step identification procedure as follows.

Step 1: apply indicator-specific deprivation cutoffs to each indicator

The indicator-specific deprivation status of household i in the k-th indicator belonging to dimension j, d_{jk}^{i}, is defined as

$$d_{jk}^{i} = \mu\left(y_{jk}^{i} < z_{jk}\right) \tag{1}$$

where $\mu(\cdot) = 1$ if expression is true; otherwise $\mu(\cdot) = 0$.

Step 2: within each dimension, aggregate deprivation of each indicator to dimensional deprivation

Denote the deprivation of household i in dimension j as s_j^i, so

$$s_j^i = \sum_{k=1}^{d_j} d_{jk}^{i} \cdot w_{jk} \tag{2}$$

Step 3: aggregate dimensional deprivation to overall deprivation

Denote the overall deprivation of household i as s^i, so

$$s^i = \sum_{j=1}^{d} s_j^i w_j \tag{3}$$

Step 4: employ the poverty cutoff to identify the poor

Poverty cutoff is denoted by k, where $0 < k < d$. P^i, the poverty status of household i, is formulated as

$$P^i = \mu\left(s^i > k\right) \tag{4}$$

Until now, we have solved the identification problem. The following procedure is to overcome the aggregation problem, the standard method of which is to calculate headcount ratio, i.e. the incidence of the poor. It can be defined by

$$H = \frac{1}{n}\sum_{i=1}^{n} P^i \tag{5}$$

However, as mentioned by Alkire and Foster (2011), one of the obvious weaknesses of this index lies in the violation of 'dimensional monotonicity'. That means, given other things, if the one who has been identified as 'poor' is deprived in one more dimension, this headcount index remains unchanged. As early as in the Sen's well-known paper on the ordinal approach to poverty measurement (Sen, 1976), monotonicity has been listed as one of the most important axioms that a valid poverty index

should satisfy. Therefore, adjusted headcount ratio denoted by M is introduced as a more reasonable and valid poverty measurement compared with headcount ratio H.

Proposed by Alkire and Foster (2011), it combines information both on the incidence of the poor, i.e. headcount ratio H, and the degree of the poverty, i.e. average deprivation share among the poor defined by

$$A = \frac{1}{d}\frac{\sum_{i=1}^{n} P^i s^i}{\sum_{i=1}^{n} P^i} \tag{6}$$

Formally, adjusted headcount ratio

$$M = HA = \frac{1}{nd}\sum_{i=1}^{n} P^i s^i \tag{7}$$

which also represents the total deprivation experienced by the poor divided by the maximum deprivation that could be experienced by the entire population.

The adjusted headcount ratio satisfies the following properties.

Subgroup decomposability: Overall poverty is the weighted average of subgroup poverty, where weights are subgroup population shares. Formally, suppose population can be divided into q groups. Let θ_g be the population share of subgroup g. Denote M_g as the adjusted headcount ratio of subgroup g, so we have

$$M = \sum_{g=1}^{q} \theta_g M_g \tag{8}$$

Dimensional decomposability[25]*:* Overall poverty is the weighted average of *censored* dimensional deprivation, where the weights are the relative across-dimension weights. "Censored" is used to stress the fact that the deprivation of those who are deprived in some indicators of a given dimension but not fall into the group of the poor are censored/not taken into account. Dimensional deprivation of a given dimension is defined as the adjusted headcount ratio of this dimension. Therefore, dimensional decomposition only holds after the poor have been identified. Let C_j be the censored dimensional deprivation of dimension j,

$$C_j = \frac{1}{n}\sum_{i=1}^{n} P^i s_j^i \tag{9}$$

Denote ω_j as the relative weight of dimension j, i.e. $\omega_j = \frac{w_j}{d}$. So the overall adjusted headcount ratio

[25] The dimensional decomposability has a significant change compared with Alkire and Foster (2011) since we introduce the indicators to the framework so as to apply the first cutoff at the indicator level rather than dimension level. However, dimensional decomposition still holds, which will be demonstrated in the appendix.

$$M = \sum_{j=1}^{d} \omega_j C_j \tag{10}$$

This formula offers a decomposition of M by dimensions and thus can shed light on the sources of poverty. The proportional contribution of deprivations in dimension p to overall poverty, denoted as R_p with $\sum_{j=1}^{d} R_j = 1$, can be measured by

$$R_p = \frac{\omega_p C_p}{M} \tag{11}$$

Therefore, the relative contribution of dimension p and $q, p, q = 1,2, \dots d$, is

$$\frac{R_p}{R_q} = \frac{\omega_p}{\omega_q} \frac{C_p}{C_q} \tag{12}$$

which has two components: the relative weights and the relative censored dimensional deprivation.

Replication invariance: Poverty is evaluated relative to the population size, i.e. comparison across different sized populations is valid.

Symmetry: Switching achievements between two analysis units do not have influence on the measured poverty. This ensures that no greater emphasis is placed on any units.

Focus: Any improvement of the non-poor cannot affect the measured poverty, which is referred as poverty focus; also the improvement in any non-deprived dimensions, no matter it is of the poor or the non-poor, cannot affect the measured poverty, which is referred as deprivation focus.

Monotonicity: Weak monotonicity ensures that poverty does not increase when there is an unambiguous improvement in achievements. Monotonicity additionally requires poverty to fall if the improvement occurs in a deprived dimension of a poor. Dimensional monotonicity specifies that poverty should fall when the improvement removes the deprivation entirely.

Non-triviality: Adjusted headcount ratio achieves at least two distinct values.

Normalisation: Adjusted headcount ratio achieves a minimum value of 0 and a maximum value of 1.

Weak transfer: It means that an averaging of achievements among the poor generates a poverty level that is less than or equal to the original poverty level.

Weak rearrangement: It means that an association decreasing arrangement among the poor generates a poverty level that is less than or equal to the original poverty level.

Two cutoffs -deprivation cutoff as well as poverty cutoff - must be chosen in order to implement this methodology. As has been pointed out by Alkire and Foster (2011a), the dual cutoffs significantly differ from one another. For the former one, it can be chosen by convention since there exists a general understanding on what a given cutoff means; while for the latter one, it seems more arbitrary

and less grounded since it works across the dimensions, where general understanding is hard to be applied. One possible method is to choose it on the basis of the specific policy goals or interests of evaluation (Alkire & Foster, 2011a), which still seems a little bit empty. To make it more detailed and easier to handle, we suggest introducing cumulative distribution function of overall deprivation to the identification process. It brings about at least two advantages. First, it provides a full picture of the overall deprivation, which should be of great help to pick up some relative meaningful poverty cutoff. Second, it makes contribution to the policy evaluation and design in the sense that it represents graphically which groups benefit from the intervention and even which group takes the largest advantage, which may direct the further quantitative study.

The empirical cumulative distribution function of deprivation score indicates the proportion of population with value of deprivation score s^i less than k. To be specific, for each possible poverty cutoff k, it estimates the proportion of population not identified as poor.

The methodology described above will be applied to the data collected from rural China to analyse the overall multidimensional poverty as well as the poverty impact of out-of-pocket health expenditure. Considering the characteristics of the data set and sampling design, dimensions, indicators, cutoffs and weighting structures involved in this study are discussed in detail in the following sections.

4. Data resources and sampling design

The data[26] used in the following analysis are extracted from 2009 Ningxia (China) Household Health Survey (NHHS) that were collected by the programme 'TB Control in China: Alignment of Health System Incentives' funded by Bill and Melinda Gates Foundation. The primary goal of this programme is to demonstrate how to enhance the effectiveness and efficiency of tuberculosis control by removing China's health system barriers. In addition, it aims to (1) improve the delivery of preventive and primary health services and people's access to them; (2) reduce financial barriers to access care for all health conditions by means of providing more generous benefit package of health insurance; (3) implement reform on provider payment to create incentives for providers to improve performance in maternal and child health, immunization and chronic disease management.

This data set is interesting for the multidimensional poverty study due to the following reasons. First, it can be used to analyse the relationship between monetary poverty and multidimensional poverty of the poor area in a developing country. The survey was conducted at five counties of rural Ningxia, one of the poorest provinces in China. In terms of per-capita net income of rural households, in 2009, the national average is around 800 US dollars while the average of Ningxia is 620 US dollars, which is ranked, from rich to poor, 22 in the total 31 provinces. Shanghai placed the top 1 with the average of

[26] Data are provided by Health Economics Research Centre, University of Oxford. For the details about the programme 'TB Control in China: Alignment of Health System Incentives', please see http://www.herc.ox.ac.uk/research/tbcontrolchina.

1900 US dollars which is more than three times compared with that of Ningxia[27]. Second, it provides detailed health expenditure information for checking and discussing the reliability and validity of health expenditure data, which is of importance for interpreting results since health expenditure serves as one of the dimensions in our multidimensional poverty analysis.

The sampling scheme is such that all the towns in the county were sampled, in each town, the villages were stratified according to their economic situation (i.e. rich, fair and poor) and then in each stratum, 40% of the villages were sampled randomly. 33 households were selected by systematic sampling in the sampled villages; for those households which were sampled but do not want to cooperate, the household before or after was taken as a replacement.

Approximately 2500 households and 11500 individuals were interviewed in 2009 52% of the respondents were male and the average age was around 30. It is noteworthy that less than 30% of the respondents completed the 9-year compulsory education, i.e. at least achieved the middle school level. The illiteracy rate of the interviewed men and women are 25% and 42%, respectively. The ought-to-be rate of completing at least middle school is around 70% since more than 70% of them were older than 15 by the day of interview. As for the occupations, nearly half of them were farmers and another one-fifth were migrant workers. The per-capital net income was about 400 US dollars[28], which is only two-third of the average of Ningxia. Given these descriptive figures based on education, occupation and income, apparently, many households in Haiyuan fell into the category of low socioeconomic status. This study is going to explore poverty on the basis of multidimensional poverty measurement proposed by Alkire and Foster (2011a) and analyse the relationship between monetary poverty and multidimensional poverty.

5. Dimensions, indicators and cutoffs

The potential dimensions that a poverty measure might reflect are quite broad such as health, education, standard of living, empowerment, work, environment, safety from violence, social relationships and culture among others (Alkire & Santos, 2010). As far as which dimensions should be considered in the research and how to choose them, Sen (2004) gave the following guidelines: (1) focus on those which are of special importance to the society or people in question; (2) focus on those which are an appropriate focus for public policy, rather than a private good or a capability which cannot be influenced from outside. (Alkire, 2008) conducted a comprehensive survey to understand how researchers selected dimensions and what dimensions they chose in practice.

[27] Data from 2010 China Statistical Yearbook complied by National Bureau of Statistics of China, published by China Statistics Press. It is available from http://www.stats.gov.cn/tjsj/ndsj/2010/indexeh.htm.
[28] Data from 2010 Ningxia Statistical Yearbook complied by National Bureau of Statistics of China - Ningxia Branch, published by China Statistics Press.

Drawn on the experience of the completed researches as well as taken the data availability into consideration, income, living standard, education and health status are selected to serve as our baseline dimensions. In addition to these four common agreed dimensions, the fifth dimension -health expenditure- is introduced to our measurement due to our particular research interest in estimating the impoverishment effects of health expenditure. For selecting indicators of each dimension, the principle of accuracy is applied, i.e. as many indicators as necessary are used given that the data are available. As far as the deprivation cutoffs are concerned, most of them are determined by convention since for many variables there is a general understanding of what a given cutoff level means and how to go about selecting it (Alkire & Foster, 2011a). For those that are hard to decide by convention, sensitivity analysis will be employed to check the robustness of the results. All the dimensions, indicators and the corresponding cutoffs are summarized in Table 1. The relevant descriptive statistics are provided in Table 2. The correlation of indicators within each dimension is attached in the appendix. For all the correlation matrices in this paper, dummy variables equal to 1 means deprived; equal to 0 means not deprived.

Table 1: Dimensions, indicators and cutoffs

Dimensions	Indicators		Deprived *if*
Disposable income	per-capita household expenditure		below 2009 national poverty line[29]
	family loans		have family loans due to food consumption
Living standards	kitchen space		kitchen is not separated
	building material		not made of brick and concrete or timber
	floor material		dirt floor
	drinking water		primary source is not tap water
	cooking fuel		typically not use electric or natural gas
	Phone		no phone
	colour TV set		no colour TV set
	durable consumer goods (bicycle, motorcycle, car, fridge, washing machine, farming machine, VCD/DVD, camera, air-conditioning)		household doesn't have at least two of the nine durable goods
Education	compulsory education of children		not all the children between 6 and 16 go to school
	educational attainment of adults		average schooling year of adults in the household is less than 6
Health status	outpatient need		have household member who reported ill in the past two weeks
	inpatient need		have household member who had inpatient need in the past one year
	diagnosed chronic disease		have household member had diagnosed chronic disease in the past half a year
Health expenditure[30]	A1	out-of-pocket expenditure share	above a given threshold[31] which is 10% of total expenditure
	A2	residual income net of out-of-pocket expenditure	below the absolute poverty line
	NA1	normative out-of-pocket expenditure share	above a given threshold which is 10% of total expenditure
	NA2	residual income net of normative out-of-pocket expenditure	below the absolute poverty line
	unmet inpatient need		have household member who had unmet inpatient need in the past one year

[29] 2009 Ningxia poverty line is 1350 RMB per-capita income. This cutoff will be substituted with (poverty line*1.2) as an indicator of 'potentially poor' in the sensitivity analysis.

[30] For the health expenditure dimension, there are four groups of indicators labelled A1, A2, NA1 and NA2, respectively, which are based on two definitions of catastrophic health expenditure from two perspectives. Only one of them is employed at every turn to calculate the poverty indices.

[31] The threshold is arbitrary and regardless of the household characteristics.

Table 2: Descriptive statistics of all the dimensions

Variables	Mean	Std. Dev.
per-capita household income below poverty line	0.079	0.270
have loans due to food consumption	0.072	0.258
kitchen is not separated	0.462	0.499
building material is not brick and concrete or timber	0.530	0.499
floor material is dirt	0.188	0.391
primary drinking water is not tap water	0.882	0.322
typically not use electric or natural gas	0.884	0.321
no mobile phone or landline	0.189	0.391
no colour TV set	0.160	0.367
the number of durable goods	2.621	1.792
- have no durable good	0.112	0.315
- have less than two durable goods	0.292	0.455
- have less than three durable goods	0.504	0.500
not all the schooling-age children go to school	0.061	0.240
average year of schooling of the adults	5.060	2.795
- less than 6	0.517	0.500
have outpatient needs	0.513	0.500
have inpatient needs	0.297	0.457
have dignosed chronic diseases	0.219	0.413
have unmet inpatient need	0.069	0.254
catastrophic health expenditure		
- budget share (positive)	0.371	0.483
- residual income (positive)	0.128	0.335
- budget share (normative)	0.580	0.494
- residual income (normative)	0.356	0.479

5.1. Disposable income

Two indicators are chosen to represent this dimension: per-capita household expenditure and family loans due to the food consumption.

Although there is a recall income question in the survey, per-capita household expenditure including the market value of the consumed self-produced goods is taken as a proxy for per-capita household income as suggested by (Morris et al., 2000), which tested the validity of proxy measures of household income using the rural Africa data. They concluded that, in the low-resource settings, household expenditure is an accepted alternative to household income even more accurate than the recalled household income. Besides the recalled income always suffers from the so-called recall measurement error, it is further more lack of reliability when the majority of respondents are self-employed farmers. In this case, some items such as the market value of the consumed self-produced goods would not be taken into account if information on income was sought but it should be considered as a part of income. Morris et al.'s findings suggest that household expenditure may be a

valid proxy for the household income in the NHHS data given the facts that they have the most important characteristics in common, which are: (1) NHHS data were collected in the rural area of Ningxia which is one of the poorest provinces in China; (2) more than half of the adult respondents reported themselves as a self-employed farmer. In order to validate this proxy in NHHS data, households are grouped into 5 quintiles according to their per-capita total household expenditure. Table 3 shows that self-produced food consumption accounts for a relatively large proportion of total food consumption no matter what economic situation they were in. The households are defined as being deprived from the perspective of income if their per-capita household expenditure[32] is below the absolute Ningxia provincial poverty line which is 1350 RMB (approximately 208 US dollars). The summary statistics shows that around 8% of the households were deprived in term of this indicator. In other words, if the traditional unidimensional measurement was employed in the poverty analysis, 8% of the respondents would be identified as the poor. Note that this provincial poverty line is lower than the $1 a day common international poverty line. If we take $1 a day as poverty line, 34.7% of the households are identified as the poor; while if we take $1.25, this figure will be 47.1%.

One more indicator –family loans due to the food consumption- is selected to represent this dimension, which is not used in standard practice. The combination of these two indicators has an advantage in further distinguishing the degree of deprivation in terms of disposable income. Around 75% of the households reported that they were holding family loans at that time, among which slightly less than 10% said that it was caused by food consumption. With these two indicators, households can be classified into three groups: extremely deprived, i.e. those who are below the cutoffs of both indicators, deprived to a certain extent, i.e. those who are below one of the two cutoffs and not deprived, i.e. those who are above the cutoffs of both indicators. The proportions of these three groups are 0.80%, 13.52% and 85.68%, respectively. In contrast, if only the poverty line indicator is applied, the proportions of the deprived and not deprived are 7.94% and 92.06%, respectively.

[32] Recall that per-capita household expenditure is used as a proxy of per-capita household annual income.

Table 3: Self-produced food consumption, food consumption an total household expenditure[33]

quintiles of per-capita household expenditure	per-capita self-produced food consumption	per-capita food consumption	per-capita total household expenditure	% self-produced food in food consumption	% food consumption in household expenditure
1	110.98	620.63	1402.34	17.88%	44.26%
2	173.54	945.04	2291.48	18.36%	41.24%
3	259.74	1195.74	3135.77	21.72%	38.13%
4	329.00	1476.34	4400.89	22.28%	33.55%
5	351.71	1774.75	10466.42	19.82%	16.96%

5.2. Living standards

Taking living standards[34] as one of the dimensions in the multidimensional poverty measurement is common agreed. The 2010 United Nations Development Programme Human Development Report introduced the Multidimensional Poverty Index proposed by Alkire and Foster[35], in which living standards were established as one of the three critical dimensions. In this UNDP report (Alkire & Santos, 2010), six indicators included to represent the living standards dimension are as follows: clean drinking water, improved sanitation, the use of clean cooking fuel, electricity, flooring material and the ownership of consumer goods. They argued that the justification for the first three indicators has been well-documented in the MDG literature[36]; the following two were used to provide some elementary indication of the quality of housing; and the last one -including radio, television, telephone, bicycle, motorbike, car, truck and refrigerator- has been adequately discussed in the literature too.

On the basis of this indicator list that has achieved certain common agreement at the international level, as well as taking NHHS data availability into account, eight indicators are selected to represent this dimension. Whether kitchen is separated or not serves as the indicator to assess the space availability. Building and floor materials can indicate the quality of dwelling. The type of drinking water and cooking fuel imply the basic dwelling-related services and last but not the least, phone, colour TV set, the number of goods owing in bicycle, motorcycle, car, fridge, washing machine, farming machine, VCD/DVD, camera, air-conditioning are employed to show the ownership of durable consumer goods. It is to be remarked that the type of the toilet, a common used indicator in

[33] The figures are in sample mean.

[34] There are two approaches to measure living standards: (1) direct measure, such as income, expenditure or consumption; (2) proxy measure, i.e. measuring living standards from information on housing characteristics and household ownership of durable goods (O'Donnell et al. 2008). However, conventionally, living standards dimension in the Alkire and Foster multidimensional poverty literature measures living standards by the proxy approach. The direct approach of measuring living standard is usually applied in the income dimension (Alkire & Santos, 2010; Foster, 2007).

[35] For more information, please see http://hdrstats.undp.org/en/indicators/38406.html.

[36] MDG stands for eight Millennium Development Goals that are end poverty and hunger, universal education, gender equality, child health, maternal health, combat HIV/AIDS, environmental sustainability and global partnership. It is a blueprint agreed to by all the world's countries and all the world's leading development institutions. They have galvanized unprecedented efforts to eradicate extreme poverty by the target date of 2015. For more information, please check http://www.un.org/millenniumgoals/bkgd.shtml.

other studies, is excluded since it cannot play a role in identifying the poor. According to the descriptive statistics, only 1% of the households had access to the flushing toilet[37], which means, almost all the households, no matter the extremely poor or the relatively rich, are all deprived in terms of having access to clean toilet. We are aware that the deprivation of living standards is generally underestimated due to this exclusion. As far as the cutoffs of each indicator are concerned, in this dimension, most of the deprivation cutoffs can be determined by convention except for the number of the durable goods. Given the fact that its mean is less than three, in the benchmark analysis, the cutoff is set as two. The descriptive statistics are summarized in Table 2. We can find that a quite high fraction of households are deprived in having access to clean drinking water and clean cooking fuel.

5.3. Education

Choosing education as one of the dimensions in multidimensional poverty measurement has reached a consensus. One of the eight Millennium Development Goals of United Nations is to achieve universal primary education. To represent education dimension, some of the studies use whether at least one household member completed primary school and whether all children of school age are attending school in the years 1 to 8 as the indicators (Alkire & Santos, 2010; Santos & Ura, 2008). Taking only one adult's education level to represent all the adults' in the household is based on the idea that all the household members could benefit from the abilities of a literate person in the household, regardless of each person's actual level of education (Basu & Foster, 1998). In the study on the multidimensional poverty of Mexican (Foster, 2007), gap-based indicators were applied to years of schooling for both the adults and the school-aged children. Different norms were set corresponding to their age from the normative perspective. As he has pointed out in the paper, it seems to be replacing 'what is' with 'what ought to be', which may reduce the potential information value from this dimension. He suggested an alternative as setting up a fixed-year standard across all adults and using the deprivation level of the parents as an estimate of the future deprivation of the children.

Based on an overall consideration of existing literatures as well as the characteristics of NHHS data, two indicators -compulsory education of children and educational attainment of adults- are selected to reflect this dimension. For the former indicator, the household is deprived if not all the children between 6 and 16 go to school, which is in accordance with the requirement in the relevance Chinese law. For the latter one, the household is deprived if the average schooling year of the adult household members is less than 6, i.e. on average, the adults in the household do not reach the primary school level. Using average schooling year of all the adult household members rather than the schooling year of one household member is to evaluate the stock of human capital for a given household size. This

[37] This is consistent with the corresponding statistics from report of National Health Services Survey in China, 2008 (Centre for Health Statistics and Information, MOH, China).

idea is from (Barro & Lee, 2010), which concluded that the stock of human capital measured by overall years of schooling has a significantly positive effect on the output.

Table 2 shows that, approximately, in 94 out of each 100 households, all the schooling-aged children were currently being educated at primary or secondary schools. However, the average schooling year of adult household members is only 5. If we set the cutoff at completing primary school level, around 70% of the households fall short of the standard of the adult educational attainment. Moreover, in this dimension, 3.47% of the households are deprived in both indicators; 50.93% of them deprived in one of the two indicators and 45.59% of them are not deprived.

5.4. Health status

Along with living standards and education which we have discusses in the previous sub-sections, health is also one of the three crucial dimensions in the 2010 UNDP Report (Alkire & Santos, 2010). In their terms, health was the most difficult dimension to measure due to the data availability problem. To be specific, comparable indicators of health for all household members are generally missing in the household surveys. They admitted that the two indicators they employed -child mortality and nutrition- are out of the domain of the ideal health indicators. Although Foster (2007) used access to healthcare as the indicator of health dimension, he argued that the correct indicators to represent the health dimension are those which are related to health status. With this respect, NHHS shows a significant superiority, because it was designed as a survey specialized in collecting health related information on household, such as self-accessed health status, outpatient and inpatient record and so on. We are therefore in the position to use some of the ideal health indicators described by Foster (2007): outpatient needs, inpatient needs and diagnosed chronic disease. We consider deprived those households which have at least one household member who reported illness in past two weeks, those with at least one household member who had inpatient need in the past one year and those with at least one member with diagnosed chronic disease Note that there was a specific question about self-assessed health status in the survey. However, we did not take it as one of the indicators since only those who answered the entire questionnaire by themselves needed to respond to the self-assessed question, which gave birth to a number of missing values (around 60% of the respondents).

Surprisingly, the descriptive statistics (Table 2) indicate that quite lots of households were facing health problems. More than half of the households had at least one member felt sick during the past two weeks from the interview date; approximately 30 out of 100 households had hospitalization needs in the past one year, while 21.9% of the households had diagnosed chronic disease patients in the past half a year. It suggests the deprivation of health may be seriously underestimated in 2010 UNDP Report (Alkire & Santos, 2010) due to representing health dimension by impropriate indicators.

5.5. Health expenditure

To the best of my knowledge, there is no published literature on multidimensional poverty taking health expenditure as a dimension or even as an indicator. However, many empirical studies on the health expenditure and poverty have shown that health expenditure contributes substantially to the impoverishment of households, especially to those from developing countries (van Doorslaer et al., 2007; Wagstaff & van Doorslaer, 2003; Berman et al., 2010; Bredenkamp et al., 2010; Limwattananon et al., 2007). Moreover, Whitehead et al., (2001) summarised the effects of excessive out-of-pocket health expenditure, referred as medical poverty trap, into four categories as follow: untreated morbidity, reduced access to care, long-term impoverishment and irrational use of drugs, all of which have significantly negative influence to the basic needs of well-being. In view of these facts, health expenditure is included in our study as one of the dimensions.

Two indicators are employed to represent this dimension, which are the unmet inpatient need and catastrophic health expenditure. The former one shows the possible distress of the household. The households are considered as being deprived if any member in the household had unmet inpatient need in the past one year; while the latter one evaluates whether out-of-pocket health expenditure excessively crowds out other non-health expenditure.

In terms of unmet inpatient needs, there were questions in the survey about whether there were inpatient needs in the past one year and whether respondents met their needs. It shows that 6.94% of the households did not meet their needs. While in terms of the catastrophic health expenditure, literatures on measuring affordability of housing and utility (Kessides et al., 2009; Thalmann, 2003) are introduced to broaden the concept of catastrophic health (Wagstaff & van Doorslaer, 2003; Xu et al., 2003). The essential idea of catastrophic health expenditure is to assess whether out-of-pocket health expenditure excessively crowds out other non-health expenditure, consequently, brings about negative influence on overall well-being, which is in accordance with the affordability concept in the literature. Hence, a household's health expenditure can be considered to be catastrophic by either the budget share –the ratio of household's out-of-pocket health expenditure and its disposable income- exceeds a certain critical point, which is the same as the conventional definition, or the residual income -income minus out-of-pocket health expenditure- is less than the standard minimum non-health expenditure. Moreover, in order to take the underutilization and overutilization of health care into account, we measure out-of-pocket health expenditure not only from positive perspective, i.e. what the observed expenditure is, but also from normative perspective, i.e. what the expenditure ought to be. Standard catastrophic health expenditure from a positive perspective is considered as benchmark. Comparisons are made between different definitions of catastrophic health expenditures as well as perspectives.

5.5.1. Measurement: budget share vs. residual income

There are two main approaches to measure affordability, which are based on budget share or residual income (Kessides et al., 2009; Thalmann, 2003). Using the same idea as Thalmann (2003) which defined housing affordability by these two approaches, health care is affordable if out-of-pocket health expenditure does not exceed a given budget share or if income minus out-of-pocket health expenditure is no less than the standard minimum non-health expenditure. In other words, we can define health care is catastrophic if out-of-pocket health expenditure exceeds a given budget share or if income minus out-of-pocket health expenditure is less than the standard minimum non-health expenditure.

Since the standard catastrophic health expenditure measurement is a relative measure, it lacks of consideration on households' ability to pay (Kessides et al., 2009). To be specific, it cannot distinguish the rich who spent 20% of his total income to pursue a high-tech treatment from the poor who spent the same proportion of the total income but still was not sufficient to meet the basic treatment need. However, the residual income approach allows us to consider "out-of-pocket health expenditure induced poverty" (Kutty, 2005; Stone, 1993).

Here is an example to illustrate how the households' deprivation statuses differ in terms of catastrophic health expenditure due to the different choice of measuring approaches (Table 4). Suppose: (1) all the households have the same disease in severity and the out-of-pocket expenditure of basic treatment is 150; (2) the minimum non-health expenditure is 1000; (3) health expenditure is 'catastrophic' if out-of-pocket share to income is larger than 10%.

Table 4: Deprivation in terms of affordability based on different measuring approaches
(Budget share cutoff = 10%, Residual income cutoff = 1000)

Households		A	B	C	D	E	F
Household income		800	1025	1100	1400	1600	1800
Out-of-pocket expenditure		50	50	150	150	150	300
Approach	Budget share	*NO*	*NO*	YES	*YES*	NO	*YES*
	Residual income	*YES*	*YES*	YES	*NO*	NO	*NO*

Household A and B are identified as 'non-deprived' by means of budget share. In fact, both can be seen as false identification but they are in different situation. Given the income of household A is below the standard non-health expenditure, in any case, it is not reasonable to identify it as 'non-deprived' after spending on health care. In terms of household B, it is considered as a false identification based on the following reasons. On the one hand, this 'non-deprived' is caused by underutilization. In other words, if they meet their basic health care needs, they will fall into the deprived group. On the other hand, even if this low out-of-pocket share is sufficient to take a basic treatment, for the poor like household B, a small out-of-pocket share that is of no consequence to the relative rich, such as household E and F, may place them in a difficult circumstance. They can meet

their basic non-health care needs only if there is no out-of-pocket health, which means they will be 'impoverished' due to this small amount spending on health care. Hence they should not be identified as 'non-deprived' in terms of out-of-pocket health expenditure. In contrast, household F is identified as 'deprived' with budget share approach. The same rationale as is discussed above, it does not consider the facts that the high out-of-pocket share results from overutilization; and for the rich like household F, a relative large proportion of out-of-pocket expenditure does not necessarily get them into trouble. However, the residual income approach has an advantage of solving this sort of misidentification problem. Household A is always deprived regardless the amount of out-of-pocket health expenditure; household B is deprived because after having paid the out-of-pocket health expenditure, they do not have sufficient budge to meet their standard non-health needs; while household F is not since they are still able to finance a minimum level of other goods' consumption. Households C, D and E represent other three possible cases in comparing these two approaches of measuring catastrophic health expenditure. It should be pointed out that the aim of the discussion above is to show that the deprivation status of a given household may differ with the applied approach rather than to conclude which approach is superior in measuring the catastrophic health expenditure.

Both of the approaches are applied in our study. Only one approach of measuring catastrophic health expenditure is considered at a time. The deprivation cutoff of budget share approach is 10% of the total expenditure. Given the arbitrariness of the threshold, sensitivity analysis will be conducted. For the catastrophic health expenditure based on the residual income approach, household is considered as being deprived if per-capita out-of-pocket health expenditure is higher than per-capita household income minus absolute 2009 Ningxia poverty line. In our benchmark case, 37% of the households are identified as being deprived; while the deprivation rate jumps to 13% if residual income approach is applied (Table 2). Table 5 shows the mean per-capita household expenditure, the mean per-capita health expenditure and two measures of catastrophic health expenditure by quintiles of per-capita household expenditure. We can find that there is a large identification gap between budget share approach and residual income approach for all the quintiles. The poorest quintile has the lowest incidence of catastrophic health expenditure by budget share approach while the highest by residual income approach, which demonstrates that, for the extremely poor, even a small share of out-of-pocket health expenditure may place them in a difficult circumstance, i.e. the income net of out-of-pocket health expenditure is not sufficient to meet their minimum non-health needs. On the contrary, the rest four quintiles show that a relative large proportion of out-of-pocket expenditure does not necessarily get the non-extremely poor into trouble. In other words, although they all have relative high incidence of catastrophic health expenditure by budget share approach[38], the majority of them still can meet their minimum non-health needs, especially for the richest three quintiles.

[38] The high incidence of catastrophic health expenditure by budget share approach is partially caused by the relative strict deprivation cutoff (10% of total expenditure).

Table 5: Catastrophic health expenditure
(budget share vs. residual income[39])

quintiles of per-capita household expenditure	per-capita household expenditure	per-capita health expenditure	catastrophic health expenditure	
			budget share	residual income
1	1402.34	133.73	29.86%	53.74%
2	2291.48	301.34	40.57%	6.90%
3	3135.77	396.12	40.79%	1.83%
4	4400.89	667.34	43.17%	2.78%
5	10466.42	1449.82	33.73%	0.98%

5.5.2. Perspective: positive vs. normative

Normative economics is that expresses value judgments about economic fairness or what the economy ought to be like or what goals of public policy ought to be, while positive economics avoids economic value judgements and is defined as the economics of 'what is' (Stone, 1993). Although most literatures studied the observed catastrophic health expenditure, i.e. 'what is', a couple of completed studies have demonstrated the necessity of considering the potential catastrophic health expenditure, i.e. 'what ought to be'. Employing Indonesia household data, Pradhan & Prescott (2002) derived a distribution of 'needed' health expenditures from the observed distribution of health expenditures, which shows that there is a gap between the observed and the 'needed' catastrophic health expenditure. Saksena et al. (2006) simulated out-of-pocket health expenditure of non-users who reported the needs of taking health care by means of Heckman-adjusted model with health status, economic and demographic variables. They also found a significant difference between the observed and the 'needed' catastrophic health expenditure using Kenya household data, which was mainly contributed by the poorer quintiles in terms of total expenditure.

The studies of Indonesia and Kenya suggest that if we only use the observed out-of-pocket health expenditure to assess the affordability of health expenditure, the incidence of catastrophe could be underestimated, especially for poorer households, because those who are too poor to afford any out-of-pocket health expenditure are excluded in the catastrophic health expenditure. Therefore, in our analysis, we not only assess the catastrophic health expenditure status of each household using the reported health expenditure, i.e. from a positive perspective, but also assess how their catastrophic health expenditure status would be if all the ill people seek standard care, i.e. from a normative perspective.

[39] The budget share cutoff is 10% of total expenditure and the minimum non-health expenditure is 2009 Ningxia local poverty line.

It is feasible to do the assessments from both the positive and normative perspectives given NHHS survey design, in which not only detailed health expenditure information was collected, but also the disease names, outpatient and inpatient needs were recorded. The aggregated out-of-pocket health expenditure data from consumption module serves as the observed health expenditure; while the 'needed' health expenditure is constructed from recorded outpatient needs and inpatient needs. It would be more precise and informative if we attach the standard expenditure to each patient in accordance with his/her disease. However, the standard expenditure of each disease is not available to date; only the average expenditure per visit for outpatient and inpatient, respectively, is available from a statistic report based on China National Health Services Survey that was conducted in 2008. Therefore, 224 RMB per outpatient visit and 3392 RMB per inpatient visit are assigned to the respondents according to their self-reported health care needs (MOH, 2008). Table 6 summarized the descriptive statistics of out-of-pocket health expenditure from positive perspective and normative perspective. With budget share approach with 10% cutoff, 58% of the households are deprived in this indicator, which is much higher than that in the benchmark case (Table 2). There are two possibilities: (1) the observed out-of-pocket health expenditure is under-reported, which is supported by the evidence of some empirical studies (Lu et al., 2009); (2) health care is significantly underutilized in the survey area, which is highly possible given the poor economic situation and empirical evidence (Pradhan & Prescott, 2002; Saksena et al., 2006)

Table 6: Descriptive statistics of out-of-pocket health expenditure (positive vs. normative)

Variable		Obs	Mean	Std. Dev.	Min	Max
positive	sample	2507	2358.76	5545.36	0	100000
	subsample[40]	2372	2493.01	5671.61	10	100000
normative	sample	2507	5884.03	6795.68	0	49620
	subsample[41]	1516	9730.39	6240.06	2446	49620

6. Weighting structure

One of the highly debatable points of multidimensional poverty measurement lies in the weighting structure, i.e. how to set the relative weights across the dimensions and the relative weights among the indicators within a dimension, which indicates the relative importance of the different deprivations (Alkire and Foster, 2011b). From a broader perspective, the selection of the dimensions and indicators are nested in the choice of weights. Specifically, while the selected dimensions and indicators are assigned positive weights, all the others excluded from the measurement are assigned nil weights. Decancq & Lugo (2010) surveyed eight approaches for setting weights in multidimensional indices,

[40] Here only consider those who reported positive out-of-pocket health expenditure in the consumption module of the survey. It indicates that 135 out of 2507 households reported zero health expenditure in the past one year.
[41] Here only consider those who reported positive out-of-pocket health expenditure in the health module of the survey. It implies that 856 out of 2372 households had spent on outpatient care before two weeks from the survey conducted.

which fall into three categories: data-driven, normative as well as hybrid. They concluded that, as long as there is no common-agreed theoretical framework on how to set weights, relying on the common sense seems to be the only choice. Therefore, in all cases, sensitivity analysis should be conducted to check the robustness of the results.

It is clear that weighting structure does matter to the analysis of multidimensional poverty. Besides sensitivity analysis, what we are able to do is to figure out the influencing mechanism of weighting structure to the given research interests, which could be of great help to interpret the results. In our case, we are interested in how the weighting structure affects: (1) the contribution of a given dimension to the overall poverty; and (2) the relative contribution across dimensions.

We assume within-dimension weights are constant in order to focus on discussing the influencing mechanism of across-dimension weights, which implies the deprivation of household i in dimension j, s_j^i, is constant (recall (2)).

From (3) and (4), poverty status of household i is

$$P^i = \mu\left(\sum_{j=1}^{d} s_j^i w_j > k\right) \tag{13}$$

From (9), (10) and (11), the proportional contribution of dimension p to overall poverty is

$$R_p = \frac{\omega_p C_p}{\sum_{j=1}^{d} \omega_j C_j} = \frac{\sum_{i=1}^{n} P^i s_p^i}{\sum_{j=1}^{d} \frac{\omega_j}{\omega_p} \sum_{i=1}^{n} P^i s_j^i} \tag{14}$$

From (9), (10) and (12), the relative contribution of dimension p and $q, p, q = 1,2, \dots d$, is

$$\frac{R_p}{R_q} = \frac{\omega_p}{\omega_q} \frac{\sum_{i=1}^{n} P^i s_p^i}{\sum_{i=1}^{n} P^i s_q^i} \tag{15}$$

Therefore, the contribution of a given dimension to overall poverty as well as the relative contribution across dimensions is highly affected by the across-dimension weights. On the one hand, relative weight itself serves as one of the determinants in both (14) and (15), which can be seen as a direct effect. On the other hand, while weighting structure affects the identification of the poor, the pattern of dimensional deprivation of the identified poor is the other determinant. This can be seen as an indirect effect.

There are two common-applied weighting structures: (1) equal weight across dimensions, i.e. the deprivation in each dimension is viewed as having equal importance; and (2) nested weights which place half the weight on income and half on the rest with equally split across the non-income dimensions, i.e. income deprivation is attached with greater importance. Taken the former one as the benchmark case, the proportional contribution of a given dimension to overall poverty and the relative contribution of two given dimensions are only affected by the pattern of dimensional deprivation of

the identified poor. If we use the latter one, comparing with the benchmark case, we more focus on the pattern of dimensional deprivation of the income-poor. Remember that the relative weights also play a role in this case. So the overall effect is ambiguous, which depends on the conjunct effect of these two determinants. In our study, both of these two weighting structures are employed. The results of equal weighting structure are reported in the results section, while the results of nested weighting structure are briefly discussed in the section of sensitivity analysis.

7. Results

We fully follow the identification and aggregation procedure described in the methodology section to measure multidimensional poverty. Also the second flow chart attached in the appendix illustrates how the methodology is implemented. The structure of this section is as follows. Firstly, we discuss the findings related to identification procedure at the household level. Then, we present multidimensional poverty indices under different settings, i.e. the findings related to aggregation procedure. Last but not the least, we report the results of subgroup and dimensional decomposition. It is worth noting the following facts: (1) all the results and findings presented in this section is based on the equal weighting structure, i.e. equal weight across dimensions as well as within each dimension; (2) in the benchmark case, out-of-pocket health expenditure is catastrophic if the "positive" expenditure exceeds 10% of total expenditure and the poverty cutoff is two.

7.1. Findings related to identification procedure

One of the most important intermediate outcomes in the identification procedure, even in the whole multidimensional poverty measurement, is the overall deprivation of each household, which integrates household's overall condition with the selection of the dimensions, indicators, deprivation cutoffs and weighting structure. For a given household, as far as the overall deprivation is derived, the poverty status can be determined by imposing poverty cutoff on the overall deprivation.

Given our study aims at measuring the impact of out-of-pocket health expenditure on multidimensional poverty and, as having discussed in the previous section, there are four possible methods to define catastrophic health expenditure, it is of importance to check whether the overall deprivation significantly varies with the definition of catastrophic health expenditure applied. Cumulative distribution curves are employed to check variation. Figure 1 is composed by four cumulative distribution curves in terms of overall deprivation: positive and normative out-of-pocket health expenditures are used to measure catastrophic health expenditure by means of budget share and residual income, respectively. It shows that there is considerable gap among these four curves, that is to say, the overall deprivation does vary with the definition of catastrophic health expenditure. Taking a step further, it implies that how to define catastrophic health expenditure measurement may affect

multidimensional poverty indices. We can have a preliminary idea about this from Figure 1. For instance, given the poverty cutoff of two, if we take residual income approach with a positive perspective, only 15% of the households are identified as poor; however, if we take the furthest one – using 10% budget share as deprivation cutoff and measure from a normative perspective- the identified poor will be around 30%, which is two times compared with the former case. Furthermore, we can compare the entire distributions of overall deprivation computed in different ways by comparing their cumulative distribution curves. For instance, a change in the measurement of catastrophic health expenditure might have different effects in different portion of the distribution, which would be missed looking only at headcount indices.

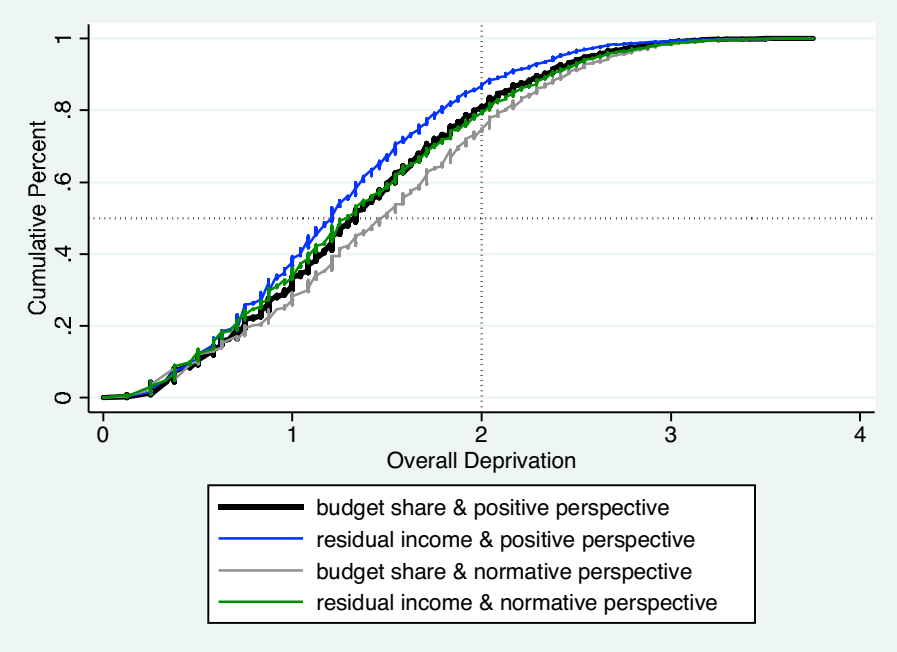


Figure 1: Cumulative distribution curves of overall deprivation

In the unidimensional poverty literature, households are identified as poor if their income or expenditure is below the selected poverty line. Overall deprivation in our methodology plays the same role as income or expenditure in the unidimensional poverty literature. Figure 2 illustrates the relationship among per-capita income, out-of-pocket health expenditure indicator and overall deprivation[42]. To interpret results correctly, we should note the following facts. Firstly, we rank each of these three variables from low value to high value, i.e. we use relative position in the population rather than absolute values in the plotting. Secondly, there are four sub-graphs due to different methods to define the indicator of out-of-pocket health expenditure: (a) positive out-of-pocket share; (b) normative out-of-pocket share; (c) income net of positive out-of-pocket health expenditure; and (d) income net of normative out-of-pocket health expenditure. Third, as we mentioned before, per-capita total expenditure is used as the proxy of per-capita income. Consequently, 'per-capita income' is

[42] In terms of the overall deprivation, deeper overall deprivation in the population associates with higher overall deprivation value.

affected by the perspective (positive or normative) of defining out-of-pocket health expenditure. Last but not the least, the relationship discussed in the following part is not conditional on whether they have health care needs.

The relationship between per-capita income and out-of-pocket health expenditure

There is no relationship between per-capita income and the share of out-of-pocket health expenditure regardless taking a positive or normative perspective. However, if we define the indicator of out-of-pocket health expenditure in terms of the residual income, we find that per-capita income net of observed out-of-pocket health expenditure is positively correlated with the per-capita income in most cases. However, in some cases, high ranking of per-capita income is associated with low ranking of residual income, the number of which increases significantly if we take a normative perspective to define out-of-pocket health expenditure.

The relationship between per-capita income and overall deprivation

The perspective of defining out-of-pocket health expenditure has indirect influence to the relationship between per-capita income and overall deprivation since, in practice, per-capita total expenditure including out-of-pocket health expenditure is used as proxy of per-capita income. However, no matter how to define out-of-pocket health expenditure, there is no clear relationship between per-capita income and overall deprivation except for those who are extremely monetary poor. Generally speaking, those who fall into the lowest 10% of per-capita income usually have high overall deprivation, not vice versa.

The relationship between out-of-pocket health expenditure and overall deprivation

The approach and perspective of defining the indicator of out-of-pocket health expenditure do matter here. The common ground of these four sub-graphs lies in that the bad performance in the out-of-pocket health expenditure is related to the high overall deprivation. But the good performance in the out-of-pocket health expenditure is not necessarily related to the low overall deprivation, which varies with the method applied to define the indicator of out-of-pocket health expenditure.

To sum up, the correlation graphs among per-capita income, out-of-pocket health expenditure and overall deprivation (Figure 2) implies that measuring poverty from unidimensional perspective or multidimensional perspective may bring about disparate results. Moreover, we should take a broader perspective to evaluate the impact of health policy rather than only measure out-of-pocket health expenditure itself.

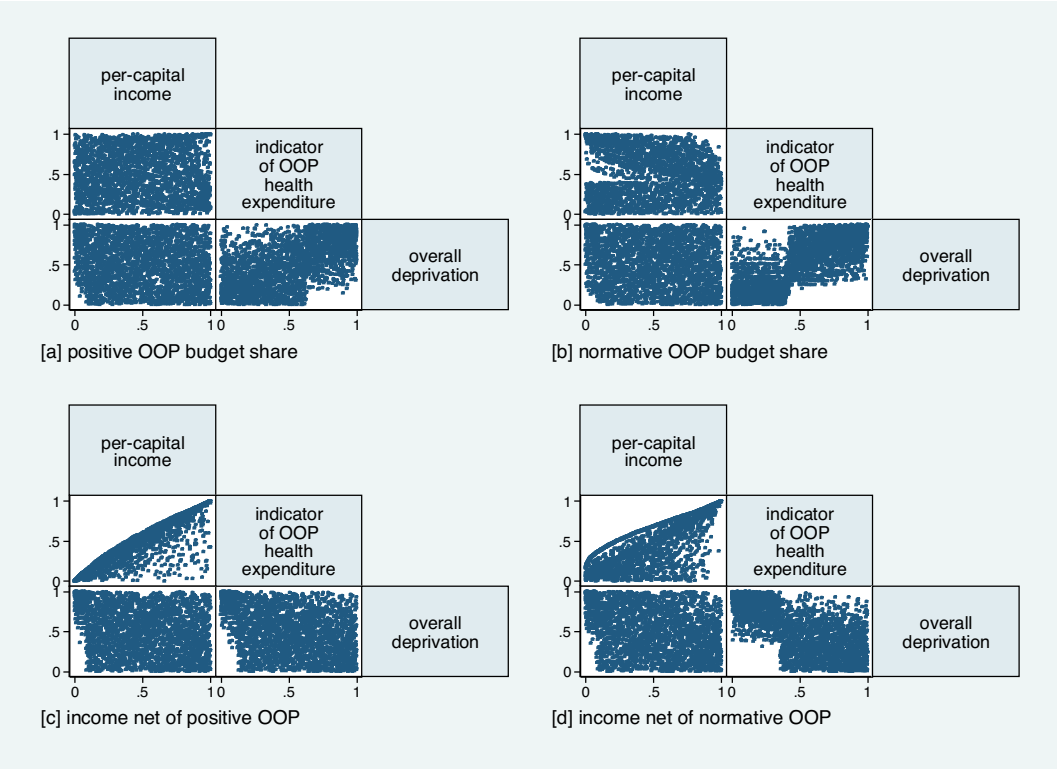


Figure 2: Relationship among per-capita income, indicator of out-of-pocket health expenditure (OOP) and overall deprivation

7.2. Findings related to aggregation procedure

After the poor are identified by the four-step identification procedure, we aggregate the information of the poor to construct multidimensional poverty indices following formula (5), (6) and (7). Under the benchmark settings, about 18.9% of the population is identified as poor, i.e. being deprived in or equivalent to two or more dimensions. Note that the headcount measure does not make a distinction between having two, three, four or five dimensions. However, the adjusted headcount ratio accounts for the number of the deprived dimensions by means of taking the average share of deprivation experienced by the poor into consideration. In this case, the average deprivation share of the poor is 48.3%, which means, on average, the poor households are deprived with respect to around half of all the indicators. Therefore, the adjusted headcount ratio - the number of deprivation among the poor in terms of the maximum number of the deprivations that can arise in the population- is 9.1%. It is obvious that if all the poor are deprived in all dimensions, i.e. average deprivation share equals to 100%, then adjusted headcount ratio equals to headcount ratio.

Table 7 summarized the multidimensional poverty indices under different settings. Keeping all the other settings as that in the benchmark case –taking poverty cutoff at 2 and measuring health expenditure from a positive perspective- except for the definition of catastrophic health expenditure switching from budget share approach to residual income approach, it shows that the headcount ratio jumps from 18.9% to 12.6%. Recall that, approximately, 36% of the households' deprivation status of

catastrophic health expenditure in terms of the budget share approach is inconsistent with that in terms of the residual income approach, in which more than 80% are identified as being deprived based on budget share approach but not by means of residual income approach. We are aware that, the higher headcount ratio with budget share approach may be caused by the relatively strict threshold that is 10% of total budget. Therefore, relative generous thresholds -15% and 20%- are also applied, which result in the lower headcount ratio of 16.3% and 14.7%, respectively. However, they are still higher compared with the one obtained from residual income approach. This implies two main possibilities: households are relatively monetary poor with a basic out-of-pocket spending (case C in Table 4) and/or, households are relatively monetary rich but overspend on health care (case E in Table 4). The government policy should be inclined to the former.

Furthermore, catastrophic health expenditure is measured from a normative perspective in order to (1) distinguish those who have catastrophic health expenditure problem due to the basic health care needs from those who fall into the group of being deprived due to the overutilization of health care; (2) take unmet needs into account to obtain the "true affordability". We find that measuring from normative perspective leads much higher poverty indices compared with that under the benchmark settings. For the headcount ratio, the former one is 25.1% while the latter one is 18.9%; for the adjusted headcount ratio, the former one is 12.2% while the latter one is 9.1%. According to these figures, it seems that the underutilization of health care should be a big concern. However, we should be highly cautious to draw any conclusion from this comparison. It is a rough estimate to some extent due to the following reasons. Firstly, the poverty indices are possibly underestimated in the benchmark case but we do not know to what extent it is underestimated. Secondly, as has been mentioned in the previous section, although we have records about the disease names, the standard or "normative" expenditure of each disease is not available, only the average expenditure of outpatient as well as inpatient is available. Hence, the "true" expenditure is actually reckoned in a very rough way. Despite these problems, measuring from a normative perspective at lease can expose the potential problems that need to be considered in the further research and policy design.

Table 7: Multidimensional poverty indices

Catastrophic health expenditure measure	Perspective	Poverty cutoff	Headcount ratio	Average deprivation share	Adjusted headcount ratio
Budget share	positive	1	0.663	0.350	0.232
		2	0.189	0.483	0.091
		3	0.009	0.644	0.006
	normative	1	0.706	0.368	0.260
		2	0.251	0.485	0.122
		3	0.014	0.648	0.009
Residual income	positive	1	0.606	0.330	0.200
		2	0.126	0.481	0.060
		3	0.006	0.663	0.004
	normative	1	0.639	0.357	0.228
		2	0.201	0.488	0.098
		3	0.012	0.651	0.008

7.3. Findings related to decomposition

One of the most important advantages or attractiveness of adjusted headcount ratio lies in its decomposability. Subgroup decomposition (formula (8)) and dimensional decomposition (formula (10)) not only can be applied respectively but also can be used conjunctly. The conjunction of two decomposition provides rich information on the patterns of poverty across a population and their differential sources (Alkire & Foster, 2011b). Table 8 presents the decomposition results -subgroup decomposition as well as dimensional decomposition- under the benchmark settings.

Subgroup decomposition is conducted based on occupation -farming/non-farming households- and ethnicity –Han/non-Han- households, respectively. For the former classification, a household falls into the group of farming households only if more than half of the adult members reported farming as their main occupation. For the latter classification, the ethnicity of households was recorded in our data.. The descriptive statistics of these two subgroups are summarized in appendix. Non-farming households account for one-third in the population, which are slightly less deprived in most indicators of multidimensional poverty measurement except one indicator - family loans- representing disposable income dimension and three indicators -kitchen, building and phone- representing living standard dimension. Han-ethnicity accounts for one-third in the population. It is at an advantage in terms of most indicators especially in education and health-related dimensions. In terms of the multidimensional poverty, there is no significant difference between farming and non-farming households. The adjusted headcount ratio is 9.6% for the former group and only 1.4% lower for the latter group. However, there does exist a multidimensional poverty gap between Han ethnicity and

other ethnicities. Compared with the Han ethnicity, the adjusted headcount ratio is doubled with regard to the non-Han ethnicities.

Decomposition among dimensions is of help to understand the contribution of each dimension to overall poverty. We find that it has nearly the same pattern of contribution among population, farming/non-farming groups and Han/nom-Han ethnicities. With the equal weights among dimensions, income dimension has the least contribution to the overall poverty, which is around 6% in all the cases we have examined; while health status dimension has the greatest contribution that accounts for around 30%, approximately. The rest three dimensions more or less contribute the same share. In other words, for all the groups, about half of the overall poverty is caused by health-related deprivations. It is worthwhile to note that the findings above are based on the equal weighting structure. Nested weighting structure will be discussed in the sensitivity analysis..

Figure 3 presents both the absolute and relative contribution of each dimension to the overall poverty. We can find that, on average, the non-farming households are only slightly better-off in terms of the multidimensional poverty while the Han-ethnic households are much better-off. Moreover, compared with other groups, multidimensional poverty of Han ethnicity is slightly more contributed by health expenditure dimension. Recall that all of the decomposition is conducted under the benchmark settings, which means catastrophic health expenditure is measured by budget share approach with the deprivation cutoff of actual out-of-pocket health expenditure exceeding 10% total expenditure. Sensitivity analysis will be conducted with more generous deprivation cutoff.

Table 8: Multidimensional poverty indices - decomposition under benchmark settings

	Population	Farming		Han	
		no	yes	no	yes
Adjusted headcount ratio	**0.091**	0.082	0.096	0.111	0.053
Subgroup weight	-	32.19%	67.81%	66.53%	33.47%
Income	5.95%	5.78%	6.03%	6.02%	5.67%
Living standard	23.36%	23.86%	23.16%	23.80%	21.56%
Education	21.19%	21.28%	21.15%	21.64%	19.29%
Health status	27.26%	27.05%	27.34%	26.90%	28.74%
Health expenditure	22.24%	22.04%	22.32%	21.64%	24.74%

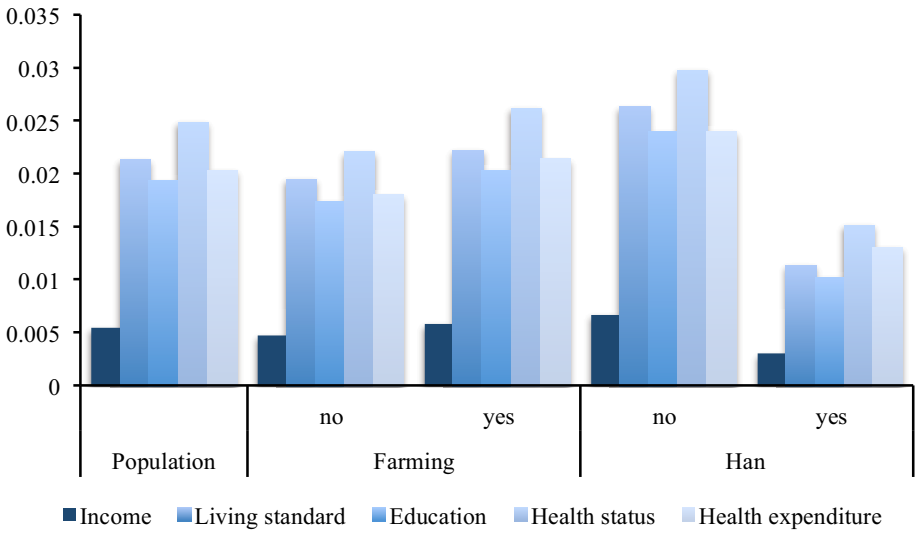


Figure 3: Subgroup and dimensional decomposition (benchmark settings)

8. A discussion about aggregated and disaggregated health expenditure data

Studies on the reliability and validity of total expenditure data are of highly importance. The completed studies comparing the differences between aggregated recall data and several-category disaggregated recall data mainly focus on the food consumption (Ahmed et al., 2006; Battistin et al., 2003; Browning et al., 2003; Gibson, 2002). By employing the data set providing food consumption information by means of both aggregated recall over the past four weeks and the diary over the past two weeks from the same household, Ahmed et al. (2006) concluded that there were substantial measurement errors existing in the one-item recall food consumption data, i.e. respondents are likely to round off the true measure causing abnormal concentrations of values in the empirical distribution. However, the diary data may not as accurate as the literature suggested. One of the problems associated with the diary data is so-called infrequency of purchase (Blisard & Blaylock, 1993; Sanchis-llopis, 2001), i.e. no purchase has been made during the short recall period although the household is a regular consumer of the goods.

As far as health expenditure data are concerned, although they are crucial for health policy design and evaluation, there has not been a sufficient concern in the health economics literatures to study the reliability and validity of health expenditure data. In other words, whether health expenditure data considerably differ due to the recall period and the number of expenditure categories and if so, whether and how they influence the empirical outcome. Lu et al. (2009), the only published research paper about the effect of survey design to date, studied how out-of-pocket and catastrophic health spending can be affected by the choice of the number of expenditure categories and recall period.

They conclude that, in most countries, fewer expenditure categories recall brought about a lower estimate of average health spending and a shorter recall period leaded to a larger estimate.

One of the main reasons of not considering the effect of survey design lies in the fact that, on most occasions, the detailed health expenditure data are not collected in practice due to the response load. However, NHHS seems to be one of the exceptions. Firstly, in the health module, it asks, individually, the respondents to record outpatient and drug expenditure in a recall period of two weeks, inpatient expenditure in a recall period of one year as well as chronic disease expenditure in a recall period of three months. And then, in the consumption module, it recalls household's total out-of-pocket health expenditure over the past one year by one question. Therefore, this survey does provide an ideal opportunity to discuss the effect of recall period and numbers of recall categories in collecting health expenditure data, which is of significantly importance given the findings on the research of food consumption data as well as Lu et al. (2009).

The basic comparison between aggregated data from consumption module and disaggregated data from health module are shown in Figure 4 and Figure 5. The first sub-graph of Figure 4 confirms the measurement error problem does present in the aggregated data in the sense that the reported out-of-pocket health expenditure is concentrated at 500, 1000, 2000 etc. However, the second sub-graph of Figure 4 shows that there are still some abnormal concentrated points in the case of disaggregated out-of-pocket health expenditure. The cumulative distribution of out-of-pocket health expenditure share is illustrated in Figure 5, which indicates that more than 50% of the households had nil out-of-pocket health expenditure in terms of the disaggregated data while it is about 5% in terms of the aggregated data. The reason lies in the fact that the out-of-pocket expenditure is composed by outpatient expenditure (including expenditure in drug store) with a recall period of two weeks and inpatient expenditure with a recall period of one year. The problems are caused by "infrequency of purchase" of outpatient expenditure (including expenditure in drug store). Those who did not receive any outpatient treatment or buy drug in the past two weeks are recorded as nil although they may spend on health care in the past one year; while those who reported outpatient expenditure in the past two weeks are extend to 26-fold to represent the whole year although it may be the only spending in the whole year. That is to say, the disaggregated out-of-pocket health expenditure is more concentrated to fewer households, which implies that catastrophic health expenditure may be overestimated in those who had positive outpatient (including drugs) expenditure record but underestimated in the rest population. Therefore, the overall effect is ambiguous. However, according to Lu et al. (2009), fewer expenditure recall categories give birth to a lower estimate of average health spending and a shorter recall period lead to a larger estimate. In our case, the aggregated out-of-pocket health expenditure has a long recall period of one year. That is to say, recall period and numbers of recall categories affect the catastrophic health expenditure in the same direction. The combination of these two effects results in an underestimated catastrophic health expenditure.

Thus, from the discussion above we could conclude that, in our case, aggregated health expenditure data are more appropriate for our analysis rather than disaggregated data. However, it suffers measurement error problems and may be underestimated to some extent, which should be kept in mind when we interpret results.

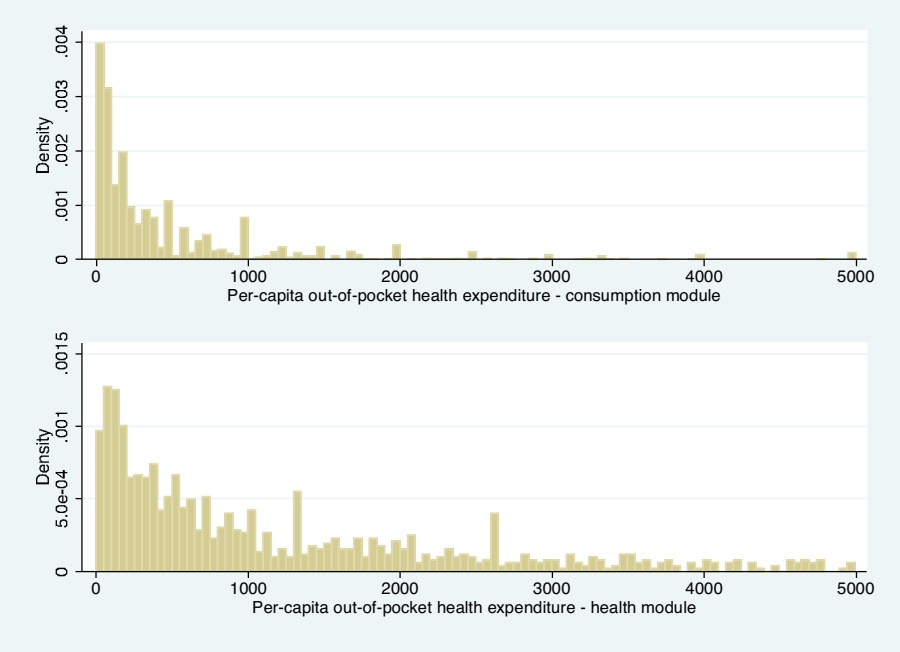


Figure 4: Histogram of per-capita out-of-pocket health expenditure

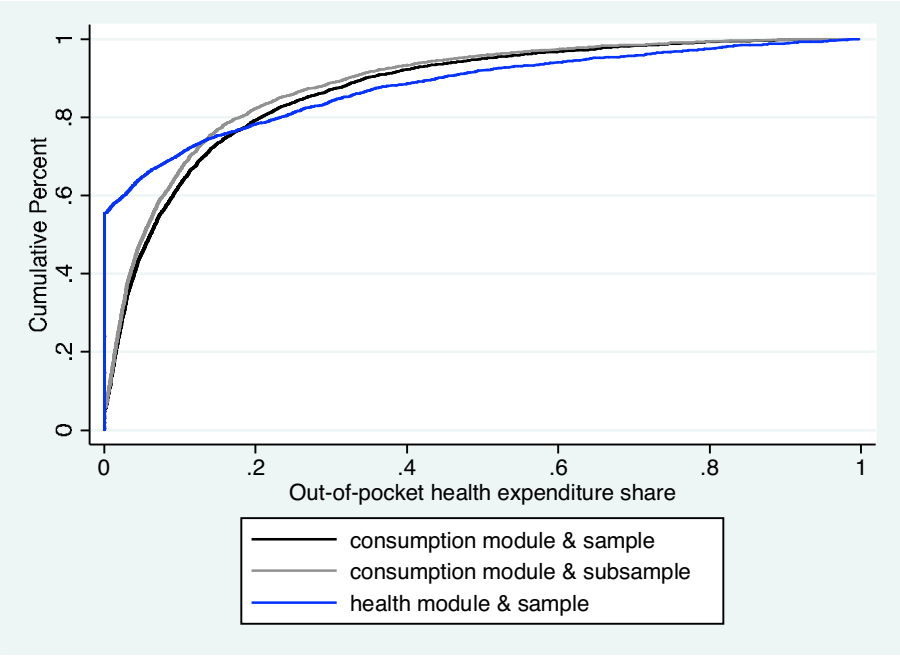


Figure 5: Cumulative distribution of out-of-pocket health expenditure

9. Sensitivity analysis

Assuming equal weighting structure, we consider two types of robustness checks: in the first one the poverty line increases by 20%, in the second, the budget share for health expenditure above which a households is considered to be deprived is raised to 20%. While the latter change affect only the *health expenditure* dimension of the multidimensional poverty index, the variation in the poverty line affects both the income poverty dimension and the health expenditure dimension when the residual income approach is used. The results are described in Figure 6 (a) and (b). We can find that, the increase of poverty line slightly has limited influence; while the increase of the budget share of catastrophic health expenditure significantly decreases the overall deprivation level but the relative contribution to the overall poverty is not significantly changed. Therefore, our findings with regards to the contribution of out-of-pocket health expenditure to the overall poverty are robust under the equal weighting structure.

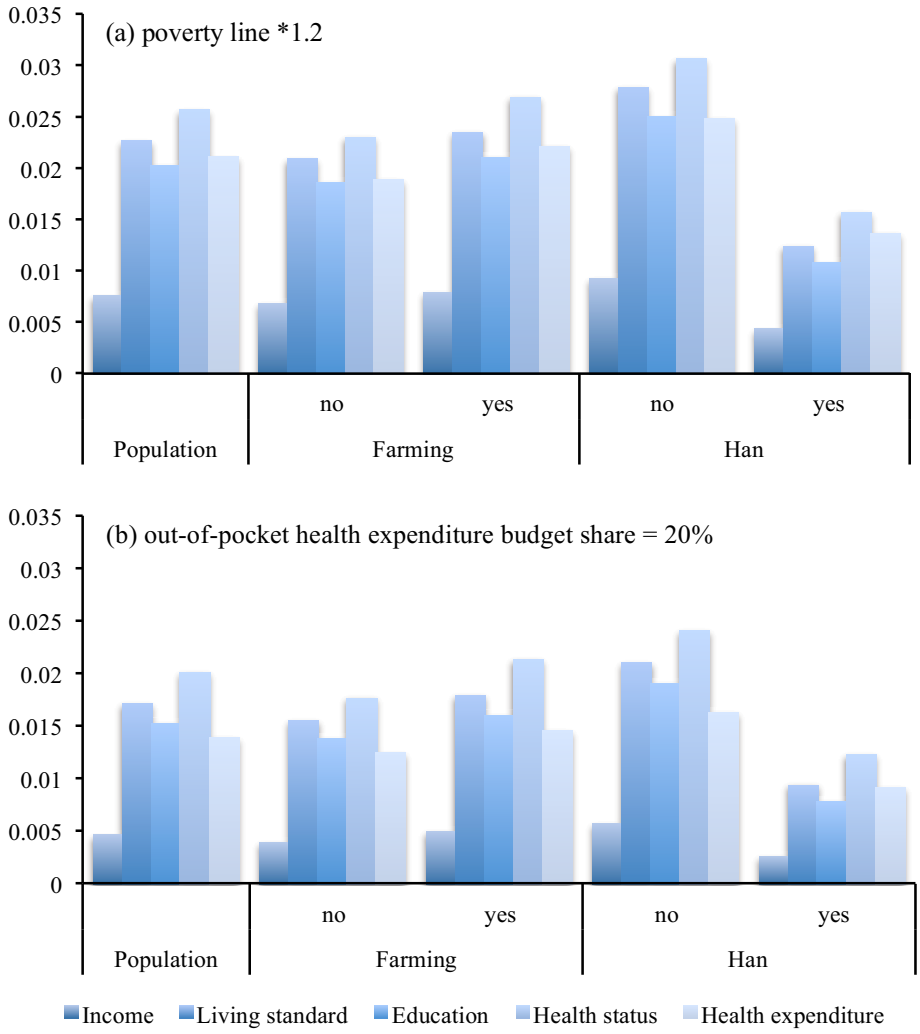


Figure 6: Subgroup and dimensional decomposition under equal weighting structure

Up to now, all the results and sensitivity analysis are conducted based on the equal weighting structure. As having discussed in the section of weighting structure, the contribution of a given dimension to overall poverty as well as the relative contribution across dimensions could be highly affected by the across-dimension weights. Recall formula (14) and (15), both of them consist of two determinants: relative weights and the pattern of dimensional deprivation of the identified poor. While the relative weights serves as a multiplier, the weighting structure has indirect but significant influence on the dimensional decomposition since different households are identified as the poor, which may have different pattern of dimensional deprivation. The first two columns of Figure 7 compare dimensional deprivation of the identified poor with equal weights and nested weights, which excludes the multiplier effects of the relative weights. It demonstrates that the relative contribution of dimensions varies significantly with the applied weighting structure. In our case, as far as employing nested weights, income dimension plays a much more important role in the overall poverty than when the equal weights are applied. It is reasonable since the nested weighting structure highly focuses on capturing the dimensional deprivation of the monetary poor while the equal weighting structure captures more households suffering deprivation in non-monetary dimensions rather than monetary dimension[43]. Moreover, it appears that the health expenditure dimension contributes least to the overall poverty, which does not imply they are deprived less in health expenditure dimension. Note that our analysis is based on the positive budget share of out-of-pocket health expenditure. For those extremely income poor, they possibly underutilize health care. Furthermore, the absolute value of dimensional deprivation of the poor is much lower in the nested weighting structure than that in the equal weighting structure mainly due to the lower incidence of the poor in the former case.

Suppose the nested weighting structure is applied. The second and third columns show the results of sensitivity analysis with higher poverty line. After the poverty line increases 20%, the absolute values of dimensional deprivation of the poor is higher due to the higher incidence of the identified poor, while the relative contribution remains unchanged.

From the discussion above, we could conclude that, given the selected weighting structure, our results with regards to the contribution of a given dimension to overall poverty as well as relative contribution compared to other dimensions are robust.

[43] Recall the deprivation rate of per-capita income and having loans due to food consumption is 7.9% and 7.2%, respectively.

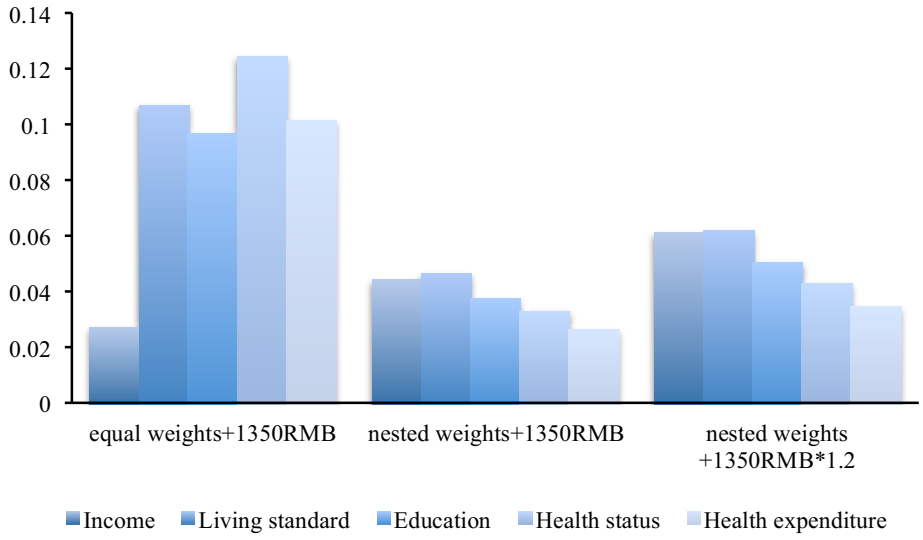


Figure 7: Dimensional decomposition (nested weighting structure)

10. Concluding remarks

We employ 2009 Ningxia Household Health Survey data to analyse multidimensional poverty with regard to a government-identified "poor" area in a developing country. After making comparison among main multidimensional poverty measurement, the one proposed by Alkire and Foster (2011a) rooting in the literature of axiomatic poverty measurement is applied in our study to measure multidimensional poverty with a set of ordinal variables.

Based on Alkire and Foster (2011a), catastrophic health expenditure literature is merged into the multidimensional poverty analysis in order to estimate the influence of health expenditure, which has not attracted much attention in the current multidimensional poverty research, on the overall multidimensional poverty. Budget share approach as well as residual income approach is applied, respectively, to measure catastrophic health expenditure. Keeping all the other settings the same, we find the proportion of households identified as multidimensionally poor by means of budget share approach is inconsistent with that obtained from residual income approach. This identification gap may come from three recourses: extremely monetary poor and underutilize health care, relatively monetary poor with a basic out-of-pocket spending and relatively monetary rich but overspend on health care. This also suggests the necessity of combining budget share approach – a relative measurement - with residual income approach – an absolute measurement - in the catastrophic health expenditure study rather than only relying on budget share approach. It is of some relevance to policy evaluation related to financial protection in the sense that it suggests a more accurate way to identify those who suffer catastrophic health expenditure. In other words, it implies the necessity of being

cautious to the results of impact evaluation that takes traditional catastrophic health expenditure as the outcome variable. To be specific, some of the results may be biased due to the inappropriate outcome measurement. Moreover, catastrophic health expenditure is measured from a normative perspective to take the underutilization and overutilization problem into account. We are aware of the roughness of this normative measurement. However, taking a step back, it addresses the potential problems that need to be considered in the further research.

Taking advantage of NHHS survey design, we discuss the data limitations of the aggregated health expenditure data from consumption module and the disaggregated one from health module. Despite that recall measurement error is well documented in the literature, we argue the data from the former are more appropriate than the latter in the context of poverty measurement.

Furthermore, decomposition among dimensions as well as groups is conducted thanks to the property of selected measurement. Assuming the equal weights among dimensions, and that the government poverty line is realistic, income dimension has the least and very limited contribution, which means that the poor are more deprived due to non-monetary deprivation than monetary poverty. However, health expenditure dimension contribute a large share to the overall poverty, which implies the significance of taking health expenditure into account. With regard to subgroup decomposition, it is found that, on average, the non-farming households are only slightly better-off while the Han-ethnic households are much better-off. It suggests that more policy that is in favor of the non-Han ethnicity should be carried out. Government should not only give priority to the development of economy but also promote the quality of education and health care, which jointly boost the well-being of non-Han ethinicity.

We are aware of the importance of weighting structure in the multidimensional poverty framework. Sensitivity analysis using nested weighting structure was conducted. The results indicate that, as far as employing nested weights which focuses on capturing the information of the monetary poor, income dimension plays a much more important role in the overall poverty than when the equal weights are applied; however, the contribution of other dimensions is still large enough to be reckoned with. We argue that, given the selected weighting structure, results with regards to the contribution of a given dimension to overall poverty as well as relative contribution compared to other dimensions are robust. The selection of weighting structure is more a policy-oriented issue which depends on the social welfare function of the policy maker.

To sum up, our study shows that, in the rural area of developing countries, especially where health insurance system is in its infancy, households may be driven into poverty by health-related deprivation more than monetary deprivation. Therefore, policy-makers should evaluate and design welfare policy from a broader perspective other than only focusing on addressing the monetary poverty. The detailed results should be interpreted cautiously in that it shares most problems of Alkire

and Foster methodology. For instance, it does not consider the interrelationships among dimensions and indicators; it is not well grounded on how to choose the weighting structure. Furthermore, it put forward the issue of survey design to measure multidimensional poverty, which is a further issue that needs to be investigated in the future.

Appendix

A. Proof of dimensional decomposability when indicators are introduced

The flow charts of Appendix B show the differences between Alkire & Foster (2011a; 2011b) and our methodology, which are caused by introducing indicators to represent each dimension so as to shift the first cutoff from dimension level to indicator level. Note that we still use dual-cutoff method to identify the poor. Although the deprivation cutoff is shifted down to the indicator level, no additional cutoff is imposed to the dimension level. Therefore, in line with Alkire & Foster (2011b), the censored headcount ratio of k-th indicator belonging to dimension j, denoted as H_{jk}, is

$$H_{jk} = \frac{1}{n}\sum_{i=1}^{n} P^i d_{jk}^i \tag{16}$$

Within each dimension, we aggregate the censored headcount ratio by with-in dimension weights, denoted as C_j, therefore

$$C_j = \sum_{k=1}^{d_j} H_{jk} w_{jk} = \frac{1}{n}\sum_{i=1}^{n} P^i \sum_{k=1}^{d_j} d_{jk}^i w_{jk} \tag{17}$$

Substitute (2) into (17),

$$C_j = \frac{1}{n}\sum_{i=1}^{n} P^i s_j^i \tag{18}$$

Substitute (3) into (7), we have

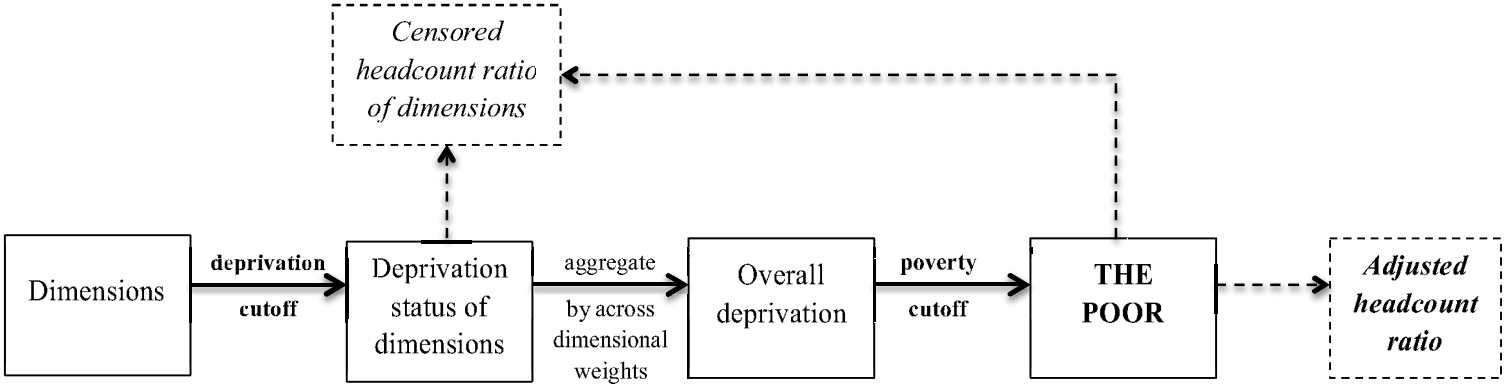
$$M = \frac{1}{nd}\sum_{i=1}^{n} P^i \sum_{j=1}^{d} s_j^i w_j = \sum_{j=1}^{d} \frac{w_j}{d} \sum_{i=1}^{n} \frac{1}{n} P^i s_j^i \tag{19}$$

Denote $\omega_j = \frac{w_j}{d}$, from (18) and (19), we have

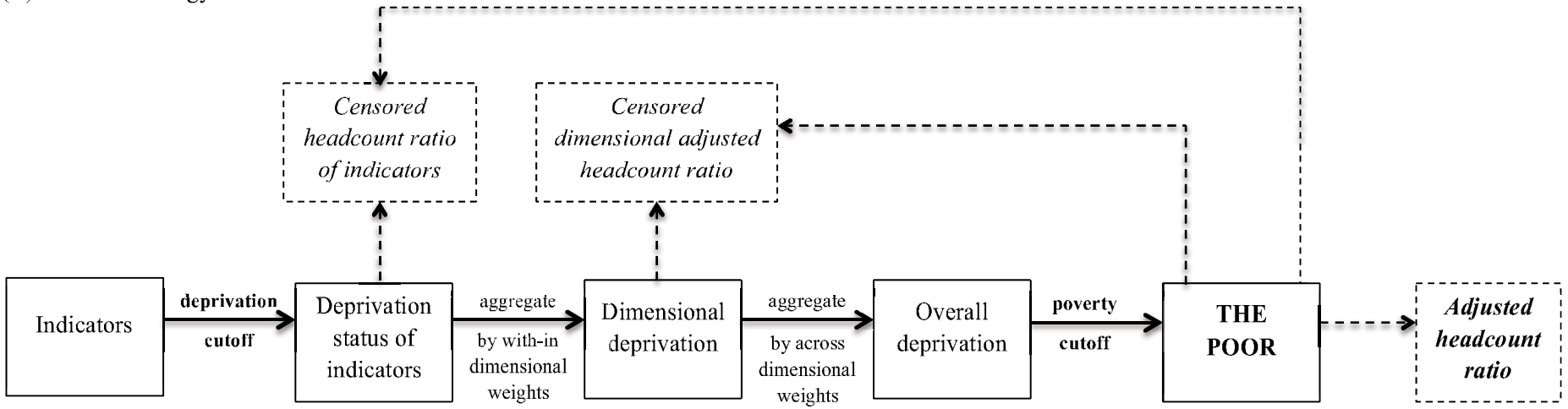
$$M = \sum_{j=1}^{d} \omega_j C_j \tag{20}$$

We define C_j as the censored dimensional deprivation of dimension j. From the proof above we can conclude that the dimensional decomposition still holds when indicators are introduced and the deprivation cutoffs are shift from dimension level to indicator level.

(I) Alkire & Foster (2011a; 2011b)



(II) Our methodology



B. Correlation matrices

Table 9: Correlation matrix of income dimension

	per-capita household income	loans due to food consumption
per-capita household income	1	
loans due to food consumption	0.033	1

Table 10: Correlation matrix of living standards dimension

	kitchen	building	Floor	water	fuel	phone	TV	durable goods
kitchen	1							
building	0.006	1						
Floor	0.110	0.161	1					
Water	-0.041	-0.021	0.030	1				
Fuel	0.037	0.087	0.029	0.060	1			
phone	0.074	0.048	0.138	-0.023	0.080	1		
TV	0.060	0.025	0.135	-0.030	0.053	0.210	1	
durable goods	0.107	0.118	0.182	0.017	0.105	0.296	0.366	1

Table 11: Correlation matrix of education dimension

	children's education	adults' education
children's education	1	
adults' education	0.024	1

Table 12: Correlation matrix of health status dimension

	no outpatient needs	no inpatient needs	no diagnosed chronic disease
no outpatient needs	1		
no inpatient needs	0.231	1	
no diagnosed chronic disease	0.191	0.147	1

Table 13: Correlation matrix of health expenditure

	have unmet inpatient need	budget share	residual income
have unmet inpatient need	1		
budget share	0.105	1	
residual income	-0.020	0.135	1

C. Descriptive statistics of subgroups

Table 14: Descriptive statistics by occupation

Variables	Non-farming (obs=807)		Farming (obs=1700)	
	Mean	Std. Dev.	Mean	Std. Dev.
per-capita household income	0.069	0.254	0.084	0.278
loans	0.082	0.274	0.067	0.250
kitchen is not separated	0.468	0.499	0.459	0.498
building material	0.544	0.498	0.524	0.500
floor material	0.166	0.372	0.199	0.399
drinking water	0.861	0.346	0.892	0.310
cooking fuel	0.860	0.347	0.895	0.307
phone	0.192	0.394	0.187	0.390
TV set	0.139	0.346	0.170	0.376
durable goods	2.656	1.877	2.604	1.751
- no	0.113	0.316	0.111	0.314
- less than two	0.292	0.455	0.292	0.455
- less than three	0.499	0.500	0.506	0.500
education of children	0.058	0.234	0.063	0.243
schooling year of adults	5.197	2.897	4.995	2.744
- less than 6	0.494	0.500	0.528	0.499
outpatient needs	0.478	0.500	0.529	0.499
inpatient needs	0.280	0.449	0.305	0.460
dignosed chronic diseases	0.197	0.398	0.229	0.420
unmet inpatient need	0.057	0.232	0.075	0.264
catastrophic health expenditure				
- budget share (positive)	0.327	0.469	0.392	0.488
- residual income (positive)	0.121	0.327	0.132	0.338
- budget share (positive)	0.544	0.498	0.596	0.491
- residual income (positive)	0.321	0.467	0.372	0.484

Table 15: Descriptive statistics by ethnicity

Variables	Non-Han (obs=1668)		Han (obs=839)	
	Mean	Std. Dev.	Mean	Std. Dev.
per-capita household income	0.076	0.265	0.086	0.280
loans	0.082	0.275	0.051	0.221
kitchen is not separated	0.544	0.498	0.300	0.459
building material	0.533	0.499	0.524	0.500
floor material	0.186	0.390	0.192	0.394
drinking water	0.871	0.336	0.906	0.292
cooking fuel	0.908	0.290	0.836	0.371
phone	0.204	0.403	0.159	0.365
TV set	0.199	0.399	0.082	0.275
durable goods	2.430	1.830	3.000	1.652
- no	0.138	0.346	0.058	0.235
- less than two	0.338	0.473	0.200	0.400
- less than three	0.558	0.497	0.397	0.490
education of children	0.076	0.264	0.033	0.180
schooling year of adults	4.353	2.605	6.467	2.623
- less than 6	0.620	0.486	0.313	0.464
outpatient needs	0.528	0.499	0.484	0.500
inpatient needs	0.318	0.466	0.255	0.436
dignosed chronic diseases	0.222	0.416	0.212	0.409
unmet inpatient need	0.079	0.270	0.050	0.218
catastrophic health expenditure				
- budget share (positive)	0.384	0.486	0.346	0.476
- residual income (positive)	0.127	0.333	0.131	0.338
- budget share (normative)	0.597	0.491	0.546	0.498
- residual income (normative)	0.370	0.483	0.328	0.470

D. Multidimensional poverty indices of subgroups

Table 16: Multidimensional poverty indices by occupation

Catastrophic health expenditure measure	Perspective	Poverty cutoff	Non-farming (obs=807)			Farming (obs=1700)		
			H	A	M	H	A	M
10% of total expenditure	positive	1	0.611	0.345	0.211	0.687	0.352	0.242
		2	0.171	0.477	0.082	0.197	0.485	0.096
		3	0.002	0.650	0.002	0.012	0.643	0.008
	normative	1	0.669	0.361	0.241	0.724	0.371	0.269
		2	0.228	0.480	0.109	0.262	0.487	0.127
		3	0.007	0.653	0.005	0.017	0.647	0.011
Residual income	positive	1	0.570	0.325	0.186	0.623	0.332	0.207
		2	0.104	0.475	0.049	0.136	0.482	0.066
		3	0.006	0.662	0.004	0.006	0.663	0.004
	normative	1	0.599	0.351	0.210	0.658	0.360	0.237
		2	0.170	0.484	0.082	0.215	0.489	0.105
		3	0.007	0.658	0.005	0.015	0.649	0.010

H: headcount ratio; A: average deprivation share; M: adjusted headcount ratio

Table 17: Multidimensional poverty indices by ethnicity

Catastrophic health expenditure measure	Perspective	Poverty cutoff	Non-Han (obs=1668)			Han (obs=839)		
			H	A	M	H	A	M
10% of total expenditure	positive	1	0.727	0.356	0.259	0.534	0.331	0.177
		2	0.228	0.484	0.111	0.110	0.479	0.053
		3	0.012	0.647	0.008	0.004	0.622	0.002
	normative	1	0.754	0.378	0.285	0.611	0.342	0.209
		2	0.300	0.487	0.146	0.154	0.475	0.073
		3	0.019	0.652	0.012	0.005	0.621	0.003
Residual income	positive	1	0.676	0.335	0.227	0.466	0.316	0.147
		2	0.150	0.484	0.073	0.076	0.467	0.036
		3	0.009	0.666	0.006	0.001	0.617	0.001
	normative	1	0.702	0.363	0.255	0.513	0.340	0.174
		2	0.236	0.492	0.116	0.130	0.474	0.062
		3	0.017	0.653	0.011	0.002	0.613	0.001

H: headcount ratio; A: average deprivation share; M: adjusted headcount ratio

Reference

Ahmed, N., Brzozowski, M., & Crossley, T. (2006). Measurement errors in recall food consumption data. *Institute for Fiscal Studies Working Papers W06/21*.

Alkire, S. (2008). *Choosing dimensions: the capability approach and multidimensional poverty*. London: Palgrave Macmillan.

Alkire, S., & Foster, James. (2011a). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7-8), 476-487.

Alkire, S., & Foster, James. (2011b). Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economics*, *9*, 289-314.

Alkire, S., & Santos, M. E. (2010). Acute multidimensional poverty : a new index for developing countries. *OPHI Working Paper*, *38*.

Atkinson, A. B. (2003). Multidimensional deprivation : contrasting social welfare and counting approaches. *Journal of Economic Inequality*, *1*, 51-65.

Barro, R., & Lee, J.-wha. (2010). A new data set of education attainment in the world, 1950-2010. *NBER Working Paper 15902*.

Basu, K., & Foster, J. (1998). On measuring literacy. *Economic Journal*, *108*, 1733-1749.

Batana, M. (2010). Aid and Poverty in Africa : Do Well-being Measures Understate the Progress ? *African Development Review*, *22*(3), 452-469.

Battistin, E., Miniaci, R., & Weber, Guglielmo. (2003). What Do We Learn from Recall Consumption Data ? *Journal of Human Resources*, *XXXVIII*(2).

Berki, S. E. (1986). A look at catastrophic medical expenses and the poor. *Health Affairs*, *5*(4), 138-145.

Berman, P., Ahuja, R. & Bhandari, L. (2010). The impoverishing effect of health care payments in India: new methodology and findings. *Economic and Political Weekly*, *45*, 65-71.

Blisard, N., & Blaylock, J. (1993). Distinguishing between market participation and infrequency of purchase models of butter demand. *American Journal of Agricultural Economics*, *75*(2), 314-320.

Bourguignon, F., & Chakravarty, Satya R. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, *1*, 25-49.

Bredenkamp, C., Mendola, M., & Gragnolati, M. (2010). Catastrophic and impoverishing effects of health expenditure: new evidence from the Western Balkans. *Health Policy and Planning*.

Browning, M., Crossley, T. F., & Weber, G. (2003). Asking consumption questions in general purpose surveys. *The Economic Journal*, *113*(F540-F567).

Cerioli, A., & Zani, S. (1990). *A fuzzy approach to the measurement of poverty*. Berlin: Springer Verlag.

Chakravarty, S. R., Mukherjee, D., & Ranade, R. R. (1998). On the family of subgroup and factor decomposable measures of multidimensional poverty. In D. J. Slottje (Ed.), *Research on Economic Inequality*. Connecticut: JAI Press.

Cheli, B., & Lemmi, A. (1995). A "Totally" fuzzy and relative approach to the multidimensional analysis of poverty. *Economic Notes by Monte dei Paschi di Siena*, *24*(1), 115-134.

Clark, D. A., & Qizilbash, M. (2005). Core poverty, basic capabilities and vagueness: An application to the South African context. *ESRC Global Poverty Research Group GPRG-WPS-026*.

Decancq, K., & Lugo, M. A. (2010). Weights in multidimensional indices of well-being : An overview. *Center for Economic Studies Discussion papers ces10.06, Katholieke Universiteit Leuven*.

van Doorslaer, E., O'Donnell, O., Rannan-Eliya, R., Ng, C. W. A. N., Pande, B. R. A. J., Tao, S., Tin, K., et al. (2007). Catastrophic payment for health care in Asia. *Health Economics*, *16*, 1159-1184.

Foster, J. E. (2007). A report on mexican multidimensional poverty measurement. *OPHI Working Paper*, *40*.

Gibson, J. (2002). Why does the Engel method work? Food demand, economies of size and household survey methods. *Oxford Bulletin of Economics and Statistics*, *64*, 341-359.

Hagenaars, A., & Vos, K. D. (1988). The definition and measurement of poverty. *Journal of Human Resources*, *23*(2), 211-221.

Haughton, J., & Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. *Handbook on Poverty and Inequality*. The World Bank.

Kessides, I., Miniaci, R., Scarpa, C., & Valbonesi, P. (2009). Toward defining and measuring the affordability of public utility services. *World Bank Policy Research Working Paper*, *4915*.

Kilpatrick, R. W. (1973). The income elasticity of the poverty line. *Review of Economics and Statistics*, *55*, 327-332.

Krishnakumar, J., & Ballon, P. (2008). Estimating basic capabilities: A structural equation model approach applied to Bolivian data. *World Development*, *36*(6), 992-1010.

Krishnakumar, Jaya. (2007). Going beyond functionings to capabilities: an econometric model to explain and estimate capabilities. *Journal of Human Development*, *8*(1), 39-63.

Krishnakumar, Jaya, & Nagar, a L. (2007). On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models. *Social Indicators Research*, *86*(3), 481-496.

Kuklys, W. (2005). *Amartya Sen's capability approach: Theoretical insights and empirical applications*. Berlin: Springer.

Kutty, N. K. (2005). A new measure of housing affordability: Estimates and analytical results. *Housing Policy Debate*, *16*, 113-142.

Limwattananon, S., Tangcharoensathien, V., & Prakongsai, P. (2007). Catastrophic and poverty impacts of health payments: results from national household surveys in Thailand. *Bulletin of the World Health Organization*, *85*, 600-606.

Lu, C., Chin, B., Li, G., & Murray, C. J. L. (2009). Limitations of methods for measuring out-of-pocket and catastrophic private health expenditures. *Bulletin of the World Health Organization, 87(3),* 238-244.

Lugo, M. A., & Maasoumi, E. (2008). Multidimensional poverty measures from an information theory perspective. *ECINEQ Working Paper*, *85*.

MOH. (2008). *An analysis report of National Health Services Survey in China*. Beijing.

Maasoumi, E. (1986). The measurement and decomposition of multi-dimensional inequality. *Econometrica*, *54*(4), 991-997.

Maasoumi, E., & Nickelsburg, G. (1988). Multivariate measures of well-being and an analysis of inequality in the Michigan data. *Journal of Business & Economic Statistics*, *6*, 327-334.

Morris, S., Carletto, C., Hoddinott, J., & Christiaensen, L. (2000). Validity of rapid estimates of household wealth and income for health surveys in rural Africa. *Journal of Epidemiology and Community Health*, *54*, 381-387.

Munda, G., & Nardo, M. (2005). Constructing consistent composite indicators : the issue of weights. *Institute for the Protection and Security of the Citizen - EUR 21834 EN*.

Nussbaum, M. C. (2001). *Women and human development: the capabilities approach*. Cambridge University Press.

Orshansky, M. (1965). Counting the poor: Another look at the poverty profile. *Social Security Bulletin*, *28*, 3-29.

Ravallion, M. (1998). Poverty lines in theory and practice: Living standards measurement study. *The World bank Working Paper No.133*.

Ruger, J. P. (2010). *Health and Social Justice* (p. 276). New York: Oxford University Press.

Sanchis-llopis, J. A. (2001). Infrequency of purchase: A model for clothing consumption with panel data. *Annales d'Economie et de Statistique*, (61), 5.

Santos, M. E., & Ura, K. (2008). Multidimensional poverty in Bhutan : estimates and policy implications. *OPHI Working Paper*, *14*.

Schokkaert, E., & van Ootegem, L. (1990). Sen's concept of the living standard applied to the Belgium unemployed. *Recherches Economiques de Louvain*, *56*, 429-450.

Sen, A. (1985). *Commodities and Capabilities*. North-Holland.

Sen, Amartya. (1976). An ordinal approach to measurement. *Econometrica*, *44*(2), 219-231.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656.

Silber, J. (2007). Measuring Poverty : Taking a multidimensional perspective. *FEDEA-Documento de trabajo*, *14*.

Stiglitz, J., Sen, A., & Fitoussi, J. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. *http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf*.

Stone, M. (1993). *Shelter poverty: New ideas on housing affordability*. Temple University Press.

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*, 103-119.

Thalmann, P. (2003). "House poor" or simply "poor"? *Journal of Housing Economics*, *12*, 291-317.

Di Tommaso, M. (2006). Measuring the well-being of children using a capability approach: An application to Indian data. *CHILD Working Paper - University of Turin*.

Townsend, P. (1979). *Poverty in the United Kingdom: a survey of household resources and standards of living*. University of California Press.

Tsui, K. (2002). Multidimensional poverty indices. *Social Choice and Welfare*, *19*, 69-93.

Vero, J., & Werquin, P. (1997). Reexamining the measurement of poverty: How do young people in the stage of being integrated in the labor force manage. *Economie et Statistique*, *8-10*, 143-156.

Wagle, U. (2005). Multidimensional poverty measurement with economic well-being, capability, and social inclusion: a case from Kathmandu, Nepal. *Journal of Human Development*, *6*(3), 301-328. doi:10.1080/14649880500287621

Wagle, U. R. (2007). Poverty in Kathmandu: What do subjective and objective economic welfare concepts suggest? *Journal of Economic Inequality*, *5*(1), 73-95.

Wagstaff, A., & van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Economics*, *12*(11), 921-34.

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

Whitehead, M., Dahlgren, G., & Evans, T. (2001). Equity and health sector reforms: can low-income countries escape the medical poverty trap? *The Lancet*, *358*(9284), 833-6.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J. L. (2003). Household catastrophic health expenditure: a multicountry analysis. *The Lancet*, *362*(9378), 111-117.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zhang, L., Cheng, X., Tolhurst, R., Tang, S., & Liu, X. (2010). How effectively can the New Cooperative Medical Scheme reduce catastrophic health expenditure for the poor and non-poor in rural China? *Tropical Medicine & International Health*, *15*(4), 468-475.

Evaluate the impact of provider payment intervention on out-of-pocket health expenditure: Evidence from a randomized social experiment in rural China

Abstract

This chapter attempts to evaluate the impacts of a block-randomized social experiment in rural China, which implemented the provider payment intervention on outpatient services. Pre-intervention and post-intervention panel data collected from household survey are employed. Both the OLS and fixed-effect difference-in-difference methods are employed to estimate the average treatment effect with a set of outcome variables related to out-of-pocket health expenditure. The estimations were exercised repeatedly using unbalanced panel, balanced panel and sub-balanced panel, respectively. We find that, although the design of the intervention is the same for two counties, the evaluation results diverse significantly. Furthermore, subgroup analysis is conducted to address the heterogeneity of treatment effect, which indicates the poor may benefit more from the interventions.

***Keywords*:** provider payment method; impact evaluation; out-of-pocket health expenditure

1. Introduction

Chinese government has budgeted a total amount of 850 billion RMB (approximately 130 billion US dollars) as the additional investment in the health care sector from 2010 to 2012 (Yip & Hsiao, 2009). Quite a portion of this money goes to subsidize the premium of a voluntary social health insurance - New Cooperative Medical Scheme (NCMS)- provided only in rural China. The aim of the additional investment on NCMS is to provide more generous benefit package to the rural residents so as to improve equal access to the basic health care as well as provide financial protection. However, both economic theory and empirical evidence have suggested that health insurance itself may have limited influence on reducing patients' financial risk, especially under the circumstance of fee-for-service provider payment system (Wagstaff & Lindelow, 2008; Yip & Hsiao, 2009), which are caused by the interaction of patients' incentive to increase health care demand (Feldstein, 1973) and providers' incentive to increase the level of demand inducement (McGuire, 2000; McGuire & Pauly, 1991). Therefore, it is of importance to explore how to utilize the government investment on health care sector in a more effective and efficient way. Many studies conducted internationally and in China have concluded that provider payment systems play a crucial role in deciding the success or failure of health care system reform (Hu et al., 2008; Yip et al., 2010).

In the context of rural China, the traditional fee-for-service method motives village doctors to over-prescribe unnecessary drugs or injections as they directly contribute to their income. Some social experiments have been done or been doing to explore a feasible provider payment method to control the quality and efficiency of providers' services. Rural Mutual Health Care (RMHC) conducted between 2002 and 2006 in Guizhou Province and Shaanxi Province is one of the succeed experiences. The intervention with regard to provider payment is to cut the link between village doctors' income and their prescription activities thoroughly. Specifically, the income of village doctors only comes from salary and a bonus based on their performance (Hsiao et al., 2008).

Another pilot study researching on the provider payment method has been working in progress from 2009 in Ningxia Province (referred to as 'Ningxia Project'). Adjusted capitation provider payment method has been introduced in the outpatient service provided by village clinics and township health centres. Contrary to RMHC which separated doctors' income from prescription activities by administrative rules, the current intervention on provider payment method keeps doctors' income linking with their prescription activities; instead, it tries to improve the quality and efficiency of doctors' service by incentive mechanism.

This chapter attempts to evaluate the impacts of the first-year provider payment intervention of Ningxia Project on the households. Two-round longitudinal household data –pre- and post-intervention - are collected both for the intervention and control groups in 2009 and 2011, respectively. Difference-in-difference estimation is applied to analyse the impacts of the intervention on household

out-of-pocket health expenditures and some other outcomes related to financial risk. It contributes to the more general literatures on the impacts of provider payment system reform in the developing countries.

The paper is organised as follows. Section 2 gives a brief overview on provider payment methods. Section 3 introduces the design of the intervention. Section 4 presents the data and descriptive statistics. Section 5 outlines the estimation methodology. Section 6 reports the results of impact evaluation, and the final section 7 concludes.

2. Overview: provider payment methods

Two provider payment methods -fee-for-service and capitation - are going to be discussed in terms of how a provider's revenue comes under alternative provider payment methods.

Fee-for-service payment method is most widely used in the health service sector, under which, a predetermined amount is charged for each discrete health service regardless the quality of the service or outcomes. There is no limit to the number of services provided per visit or the number of visits per patient (Miller 2009). So a provider's net revenue under fee-for-service method can be expressed as follows.

$$\begin{aligned} &Provider's\ net\ revenue \\ &\quad = charge\ per\ treatment \times \#treatments\ per\ patient^{44} \times \#patients \\ &\quad - \quad cost\ per\ treatment \times \#treatments\ per\ patient \times \#patients \end{aligned}$$

From the expression above, we can find that fee-for-service payment method may give birth to two incentive-related issues: (1) it gives inappropriate incentives to providers to use even over-use the high profit treatment instead of most appropriate treatment; and (2) it is lack of incentive for providers to improve the quality and efficiency of the service as they have nothing to do with their net revenue.

There is emerging evidence indicating that capitation payment method has its advantages compared to fee-for-service payment method. Under the capitation payment method, a provider (such as village doctor, township health centre) receives a single predetermined health budget for each person regardless of how sick or how well he is. Under this circumstance, providers may feel that they treat patients "with their own money". A provider's net revenue can be presented as follows.

[44] "the number of" is abbreviated to"#". #treatments per patient=#treatments per visit × #visits per patient

$$Provider's\ net\ revenue$$
$$= \overline{budget\ per\ person}^{45} \times \overline{\#person}$$
$$- cost\ per\ treatment \times \#treatments\ per\ patient \times \#patients$$

Compared with fee-for-service method, the capitation method does give a strong incentive to providers to consider the cost per patient, i.e. the cost per treatment and the number of treatments per patient needed, which is of great help to avoid the expensive and/or unnecessary treatment. Meanwhile, it may also have a positive effect on improving the quality of preventive health care service since the number of patients is highly relevant to their net revenue. However, it may over-motive providers to control the cost to some extent. Put differently, it gives providers an undesirable incentive to avoid the costly-to-treat patients or provide insufficient health service to them.

In practice, adjusted capitation payment method is commonly used in order to address the undesired problems emerging in the capitation method. The adjustments can be made in terms of the budget and/or the cost. So the adjusted capitation payment method can be expressed as follows.

$$Provider's\ net\ revenue$$
$$= F(\overline{budget\ per\ person} \times \overline{\#person})$$
$$- G(cost\ per\ treatment \times \#treatments\ per\ patient \times \#patients)$$

where $F(\cdot)$ is a predetermined function for budget adjustment while $G(\cdot)$ is a predetermined function for cost adjustment.

Under the adjusted capitation payment method, the payment that the provider can finally receive would be adjusted according to some conditions such as his quality and efficiency of the services which could be evaluated by a set of indicators. Also, in contrast with the traditional capitation method transferring all of the cost risk to the provider, the adjusted capitation method allows introducing cost risk sharing mechanism. Therefore, it could get rid of the undesired incentive - avoid the costly-to-treat patients or provide insufficient health service- if designed appropriately. To sum up, given the proper design of the budget adjustment rules and/ or cost adjustment rules, the adjusted capitation payment method could make best use of the advantages and bypass the disadvantages of traditional capitation method.

[45] $\bar{X}$ means X is predetermined.

3. Provider payment intervention in Ningxia Project

3.1. Policy introduction

The intervention has been going on in parallel in two counties in Ningxia Province –Haiyuan and Yanchi- that were selected beforehand on the basis that (1) there was no health-related pilot project running in 2009 or had taken place in the previous years; and (2) the local governments (counties and towns) were willing to involve in the social experiment. Having been aware of potential differences between these two counties, a randomized block design is applied, i.e. each county itself serves as one block and townships within each county are assigned randomly to the interventions of provider payment method. All the households live in the treatment towns are treated while all the households live in the control towns are not treated. Adjusted capitation method has been introduced into the treatment towns to replace the fee-for-service payment method since 2010. In the first year intervention, only the public health and outpatient services provided by village clinics and township health centres[46] were involved in the payment method intervention. Put differently, all the services provided by county hospitals and the inpatient services provided by township health centres are still using the fee-for-service payment method whereas their annual revenues are capped.

The budget under capitation method consists of the budget of public health and the budget of primary care (the detailed calculation is attached in the Table 1 in the appendix). However, both the budget adjustment and cost adjustment are applied. For township health centres, they receive 60-70% up front of total budget at the beginning of each year. The rest are subject to some targeted standards of services as well as performance evaluation. The former includes the cost of per service and the volume of services provided. If the total cost exceeds the total budget, the exceeding part will be partially shared only when the cost per service is up to the standard. The latter consists of a set of indicators related to the quality and efficiency of the services (see the summary of provider payment intervention in Figure 1). The provider payment method of village clinics is analogous with that of township health centres. It is worth noting that, in additional to their own public health and primary care responsibilities, township health centres are designed to involve in the managing and monitoring of affiliated village clinics as follows. First, township health centre assesses the performance of affiliated village clinics while the performance of township health centre is assessed by county health bureau. The performance of their affiliated village clinics is included in the assessment of township health centre. Second, township health centre holds the village clinics' budget and distributes it on the quarterly basis. Third, township health care shares the cost of over-run of village clinics but also shares savings.

[46] The health care delivery system in China consists of village clinics, township health centres and county hospitals in the rural areas. Village clinics and township health centres mainly provide public health and primary care services. Some of the township health centres provide limited inpatient service. Please see Yip et al. (2010) for a detailed review of China health care delivery system.

3.2. Expected impacts

We expect the provider payment intervention has direct effects on the supply side –health care providers- rather than the demand side-households. With the adjusted capitation method, a given rational provider (village clinic or township health centre[47]) maximizes his net revenue from outpatient services by choosing the optimal values of p4p score, cost per service and service volume subject to the constraints on the last two. While the first one –p4p score- is positively correlated with the net revenue; the rest are negatively correlated with the net revenue. Therefore, the expected responses of rational providers should be (1) make effort to improve the quality and efficiency of public health and primary care services such as disease prevention and management, maternal and child health care since the p4p score are determined by them; (2) control the total cost of the outpatient service but meet the requirement of cost per service and service volume, which means reducing repeated tests, over-prescription of drugs and antibiotics in outpatient services[48]. It is worth noting that, for some township health centres that are capable of providing inpatient services[49], an undesired response to the intervention may come forth, which is shifting those who can be treated with costly outpatient service to inpatient service.

Assuming that providers response to the payment intervention rationally, the intervention could have the following indirect impacts to households. First, thanks to the providers' effort of improving quality and efficiency of primary care, the utilization of outpatient services may increase, which may raise the average outpatient expenditure[50] in the short run but would reduce the total health expenditure in the long run in that people may have a lower probability to be ill or they seek care once they have minor illness. Second, conditional on outpatient utilization, average outpatient expenditure should be reduced, which results from the active cost control of providers. Third, it is possible that total health expenditure increases in the short run due to the undesired substitution effect between outpatient and inpatient services.

In order to evaluate the direct impact of the provider payment intervention, provider level data collected from health facility survey or health service information system are required. However, if only the household survey data are available, it is possible to evaluate its indirect impacts on households, which mix up two impacts: (1) how provider payment intervention (exogenous signal) affects providers' behavior; (2) how the changes of providers' behavior (endogenous signal) affect

[47] Recall that only the public health and outpatient services provided by village clinics and township health centres were involved in this intervention. For simplicity, we refer village clinics and township health centres as 'providers' in this chapter unless any explicit statement is made.

[48] In China, price scheme of health care services is highly distorted in that the labour-intensive services are priced lower than the cost whereas the capital-intensive services are priced higher than the cost. Therefore, with fee-for-service payment method, there are serious problems with regard to repeated tests, over-prescription of drugs and antibiotics so as to high out-of-pocket health expenditure (Yip et al., 2010)

[49] Village clinics are not capable of providing inpatient services at all. So this undesired substitution effect does not exist. However, some (not all) township health centres are capable of providing inpatient services.

[50] All the 'expenditure' means 'out-of-pocket expenditure' unless specified clarification is made.

households. Figure 1 summarizes the intervention design and expected impacts on both supply side and demand side.

3.3. Evaluation

Given the data availability and our research interests, we will study the impact of provider payment intervention on household health expenditures. According to the discussion above, it is highly possible to get limited or obscure results due to the following reasons. First, what we are going to evaluate is an incentive chain, the results of which depends on whether providers follow and rationally react to the designed incentive mechanism and then how households respond to the providers' change. Second, the one-year implementation period may be too short to show the clear or true impacts because providers as well as households respond to their 'new signal' sequentially not simultaneously, which takes time.

Two groups of outcomes related to household health expenditure are chosen, which are the absolute value of health expenditures and the affordability of total health expenditure. The former includes outpatient expenditure and total health expenditure. The latter includes health expenditure share in terms of total expenditure, total expenditure net of health expenditure as well as catastrophic health expenditures. In the literature, a household's health expenditure is considered to be catastrophic if the ratio between the household's out-of-pocket health expenditure and its disposable income exceeds a certain critical point. Commonly used thresholds include 30% or 40% of capacity to pay[51], or 10% of total income (Wagstaff & van Doorslaer, 2003; Xu et al., 2003). In addition, by learning from Thalmann (2003) which studied housing affordability, we suppose a household's health expenditure can be viewed as catastrophic if total expenditure net of health expenditure is below the standard minimum non-health expenditure. Both of these two methods of defining catastrophic health expenditure are applied in our analysis. Note that even if our intervention affects health expenditure share and/or total expenditure net of health expenditure significantly, it may not affect catastrophic health expenditure. Put differently, the catastrophic health expenditure can be affected only if the intervention focuses on the households nearby the threshold.

[51] "Capacity to pay" is defined as effective income net of basic subsistence needs have been met. In practice, effective income is taken to be the total consumption of the household since in many countries it is a more accurate reflection of purchasing power than income reported in household surveys; basic subsistence needs is taken to be the total consumption spending on food.

Payment method intervention

$$
\begin{aligned}
&Provider's\ net\ revenue \\
&\quad = \overline{budget\ per\ person} \times \overline{\#person} \times \alpha(p4p) \\
&\quad - cost\ per\ service \times \#services \\
&\quad - G_1(cost\ per\ service - \overline{cost\ per\ service}) \\
&\quad - G_2(\overline{\#services} - \#services) \\
&\quad + G_3(cost\ per\ service \times \#services)
\end{aligned}
$$

$$
\begin{aligned}
where \quad & \alpha(p4p)\ is\ the\ adjustment\ rate\ determined\ by\ performance\ evaluation\ p4p \\
& G_1(\cdot) = 0\ if\ cost\ per\ service \leq \overline{cost\ per\ service} \\
& G_2(\cdot) = 0\ if\ \#services \geq \overline{\#services} \\
& G_3(\cdot) = 0\ if\ cost\ per\ service > \overline{cost\ per\ service} \\
& \qquad\qquad or\ cost\ per\ service \times \#services \leq \overline{budget\ per\ person} \times \overline{\#person}
\end{aligned}
$$

Expected impacts

Provider side:

- response to p4p: improve quality and efficiency of service [R1]
- response to total cost: control total cost but meet the basic requirements on cost per service and #services provided [R2]

Demand side:

- caused by [R1]: population health improved, may indirectly reduce health expenditure but may not
- caused by [R2]: health expenditure directly reduced conditional on health utilization

Evaluation

Provider data from health facility survey or health information system

Household data from household survey

Figure 1: Summary of intervention and expected impacts

4. Data and descriptive statistics

4.1. Sampling design and data

Up to now, two-round household survey data have been collected by the Ningxia Project team in the 2009 and 2011, respectively. The 2009 data collected before any intervention was introduced serves as the baseline data. Provider payment intervention described in the previous section was introduced into the treatment towns of Haiyuan and Yanchi from the beginning of 2010. And then after one year of intervention, the project team conducted the follow-up survey to review the first-year intervention.

All the towns belonging to either Haiyuan or Yanchi were sampled. In each town, the villages were stratified according to their economic situation (i.e. rich, fair and poor) and then within each stratum, 40% of the villages were sampled randomly. 33 households from each sampled village were picked out by means of the methodatic sampling.

Around 3800 households from the two counties were interviewed in 2009. In the 2011 follow-up survey, those who could not be followed up were substituted by the households living in the same village. The follow-up rates vary from 75% to 90%[52]. Therefore, it is necessary to check that whether, on average, the new added-in households share the similar characteristics as the dropped-out households. We compare the following demographic characteristics of the households: ethnicity, highest education achievement, average age of the adults, household size and number of children. Generally speaking, there is no significant difference between the dropped-out and added-in households on these demographic variables, although, on average, the added-in households of the treatment towns seem to be slightly less educated than the dropped-out and the added-in households of the control towns are slightly younger[53].

4.2. Variables

All the definitions of key variables are listed in Table 2, the majority of which are straightforward. We will discuss wealth index and a set of outcome variables in detail.

4.2.1. Wealth index

The wealth index is used as a covariate in our analysis to account for wealth effects. It, defined as the values of possessions and living conditions of the household (O'Donnell et al. 2008), is a proxy measure of living standard on the basis of available variables describing household assets and characteristics. In practice, household ownership of durable goods (i.e. television, refrigerator, etc.),

[52] For Haiyuan, 1047 (1451) households from the treatment (control) towns were surveyed in 2009, in which 805 (1054) were successfully followed up in 2011. For Yanchi, 722 (592) households from the treatment (control) towns were survey in 2009, in which 645 (485) were followed up in 2011.

[53] The descriptive statistics also show that the added-in households have smaller household size and fewer children on average. However, the followed-up sample has the same tendency.

dwelling characteristics (such as source of water, cooking fuel, etc.) and land ownership are included. These variables are weighted to create the wealth index (Filmer and Pritchett 2001). In our analysis, principal components analysis – statistical procedures which summarize the variability among a set of variables – is employed to determine the weights[54][55].

The wealth index A_i is then given by

$$A_i = \sum_k \left[f_k \frac{(a_{ik} - \bar{a}_k)}{s_k} \right] \quad (1)$$

where, a_{ik} is the value of asset variable k for individual i, $\bar{a}_k$ its mean value, s_k its standard deviation, and f_k the weights of the first principal component (O'Donnell et al., 2008).

In terms of the interpretation to the results of principal components analysis, higher socioeconomic status is predicted by positive factor scores while the opposite is true of negative factor scores. All the variables included in as well as the results from principal components analysis is presented in Table 3.

4.2.2. Outpatient health expenditure

We have most interests in evaluating whether the intervention has affected household outpatient expenditure since the intervention mainly focused on the outpatient service. Given the data availability, there are two ways to measure household outpatient expenditure as follows.

The first method is, within each household, aggregate the individual-specific two-week outpatient expenditures (including drugs) recorded in the health module. Assuming there are n individuals living in household i, the two-week outpatient expenditure of household i, $op\_exp\_2w_i$, can be expressed as

$$op\_exp\_2w_i = \sum_{j=1}^{n} (drug_{ij} + B14.1_{ij} + B20.1_{ij}) \quad (2)$$

where $drug_{ij} = B10.1_{ij}$[56] if in year 2009; $drug_{ij} = B10.12_{ij} + B10.24_{ij}$ if in year 2011 and answer (1) to question B10.23; $drug_{ij} = B10.12_{ij} + B10.22_{ij}$ if in year 2011 and answer (2) to question B10.23.

The second method is, aggregate the individual-specific annual inpatient expenditure recorded in the health module within each household and then subtract it from the household total health expenditure recorded in the consumption module. The annual household outpatient expenditure of household i, $op\_exp\_sub_i$, can be expressed as

[54] There are three commonly used weighting schemes as follows: (1) assign equal weights to all assets; (2) assign weights determined by the asset prices; (3) assign weights determined by principal components analysis (O'Donnell et al., 2008; Filmer and Pritchett, 2001).

[55] For households that had missing values the following strategy was adopted: (1) for durable goods and specific house characteristics, missing values were coded as not owning the item or not having that house characteristic as less than 0.1% of households had missing values; (2) for house area, missing values were replaced with the mean (in 0.7% of households).

[56] $B10.1_{ij}$ denotes the answer to question B10.1 of household survey (see Figure 2) from individual j of household i. Similarly hereinafter.

$$op\_exp\_sub_i = J14.4_i - \sum_{j=1}^{n} IP_{ij} \quad (3)$$

where $IP_{ij} = C4.9_{ij} - C4.9.1_{ij}$ if answer (1) to question C4.8.1; $IP_{ij} = C4.10_{ij}$ if answer (2) to question C4.8.1.

Both of them can be criticized from various perspectives. The former suffers at least three problems. The first one is caused by the proxy response - it did not require each individual to answer the survey personally[57]. This is problematic especially to the record of outpatient expenditure (including drugs) in the sense that the respondent may not know the expenditure of other family members as it is not a "big event". The second one is due to the short recall period so as to the related "infrequency of purchase" (Blisard & Blaylock, 1993; Sanchis-llopis, 2001). The third one is caused by the design of the 2009 questionnaire. The drug expenditure in 2009 may be overestimated since the drug purchase from pharmacy and health facility was not separated at that time. But the purchase from the latter could be reimbursed by NCMS[58]. Although the proxy response is less problematic in the latter since the hospitalization is a "big event" with low incidence, it is challenged by the other two issues. First, the total health expenditure suffers measurement errors, i.e. respondents are likely to underestimate the one-item total health expenditure (Lu et al. 2009) and round off the true measure causing abnormal concentrations of values in the empirical distribution (Ahmed et al. 2006). Second, we find more than 10% of the households are assigned negative outpatient expenditure due to the subtraction.

We studied the probability of having non-negative outpatient expenditure by means of the linear probability model (Table 4). The results indicate that, for both counties, household size and average age of the adults do have significantly negative effect on the probability of having non-negative outpatient expenditure; for Haiyuan, the Han-ethnic households also have significantly lower probability to get non-sense values[59]. However, there is no difference between pre- and post-intervention. And it does not depend on household wealth or education achievement[60].

Two alternatives are used to cope with this negative data problem: (1) drop the households with negative outpatient expenditure and only estimate the non-negative subsample[61]; (2) impute the negative outpatient expenditure with the information of non-negative households[62].

---

[57] The proxy response rate is around 60% for the 2009 survey and 64% for the 2011 survey.

[58] This problem will be shown if we compare the questions of 2009 and 2011.

[59] More than 90% of the households in Yanchi are Han-ethnic.

[60] The accuracy of proxy may highly depend on the individual characteristics of the respondent rather than the household's. However, the data are not available on who was the respondent.

[61] We are aware that dropping households with negative outpatient expenditure brings about sample bias problem. However, keeping them generates the concern about data reliability and validity since they were not able to provide consistent information on health expenditure.

[62] We did imputation separately for treatment towns of Haiyuan in 2009, treatment towns of Haiyuan in 2011, control towns of Haiyuan in 2009, control towns of Haiyuan in 2011, treatment towns of Yanchi in 2009, treatment towns of Yanchi in 2011, control towns of Yanchi in 2009, control towns of Yanchi in 2011. Stata command 'uvis' is employed to do the

4.2.3. Total health expenditure

There are two possible ways to measure total health expenditure: (1) take the total health expenditure from household consumption module (J14.4); (2) construct it from the individual health module. Let the_c_i denote the constructed total health expenditure, we have

$$the\_c_i = op\_exp\_2w_i \times 26 + \sum_{j=1}^{n} IP_{ij} \quad (4)$$

It has been well documented in the literature that the one-item recalled total expenditure usually suffers measurement errors. We did check this problem in our data and found that measurement error problem presents in both cases[63]. According to Lu et al. (2009), in most countries, fewer expenditure categories recall brings about a lower estimate of average health spending and a shorter recall period leads to a larger estimate. Therefore, in our case, the total health expenditure from household consumption module (J14.4) underestimates the average health expenditure whereas the constructed one overestimates it. Since the constructed total health expenditure is not superior to the one-item recalled one, we take the latter for simplicity.

4.2.4. Affordability-related outcome variables

Affordability-related outcome variables include (1) health expenditure share in terms of total expenditure; (2) total expenditure net of health expenditure; (3) dummy variable which equals to 1 if health expenditure share in terms of total expenditure exceeds 10%; (4) dummy variable which equals to 1 if total expenditure net of health expenditure is below the local poverty line (1350 RMB). Note that all the 'health expenditure' in the group of affordability-related outcomes means the total health expenditure from household consumption module (J14.4).

4.3. Samples

Three samples from the same data source are employed in our analysis, which are referred to as unbalanced panel, balanced panel and sub-balanced panel. Unbalanced panel contains all of the two-wave data with 7624 households. Balanced panel excludes those who were interviewed in 2009 but dropped out in 2011 as well as those who were added in as substitutions in 2011, which contains 5978 households. Sub-balanced panel only includes those who have positive outpatient expenditure with the subtraction method, i.e. $op\_exp\_sub \geq 0$, which have 5320 households.

All the relevant descriptive statistics in terms of different samples are attached in the appendix (Table 5 and Table 6).

---

imputation. It imputes a single variable on a set of predictor variables by OLS regression. Predictor variables used in our imputation are chronic disease dummy, household size, ethnicity dummy, average age of the adult and per-capita food consumption.

[63] The total health expenditure from household consumption module (J14.4) is concentrated at 500, 1000, 2000 etc. However, there are some abnormal concentrated points in the constructed health expenditure too, such as 1300, 2500.

Self-purchased Drugs in the past fourteen days (individual health module):

B10.1 If the medicine was purchased at the pharmacy or at a health facility, how much did you spend on pharmaceutical products? *[2009]*

B10.12 If the medicine was purchased at the pharmacy, how much did you spend on pharmaceutical products? *[2011]*

B10.22 If the medicine was purchased at the health facility, how much did you spend on pharmaceutical products? *[2011]*

B10.23 Did you receive reimbursement from NCMS? (1) Received (2) Did not receive *[2011]*

B10.24 After reimbursement, how much did you actually pay out of pocket? *[2011]*

Visiting doctor in the past fourteen days (individual health module):

B14.1 (First visit to the doctor) How much of that was paid in cash by you?

B20.1 (Second visit to the doctor) How much of that was paid in cash by you?

Hospitalization in the past one year (individual health module):

C4.8.1 If you were enrolled in NCMS, how were you reimbursed for this hospitalization? (1) Paid the entirety up front, then went to the NCMS Office to apply for reimbursement. (2) Hospital directly lowered the hospitalization fee. (Skip to C4.10)

C4.9 During this hospitalization, how much did you spend up front? (Does not include transportation costs, personal nurse, bribes)

C4.9.1 How much did the NCMS Office reimburse you for this hospitalization? (Skip to C4.11)

C4.10 During this hospitalization, how much did you spend out of pocket? (Does not include transportation costs, personal nurse, bribes)

Total health expenditure (household consumption module)

J14.4 Expenditure on the following commodities over the past year (12 months): Medicine, medical expenses, etc. (only includes what you personally paid for health services and medicine)

Figure 2: Household survey questionnaire extracts

5. Estimation strategy

5.1. Average treatment effect

To evaluate the causal effect of an intervention, conducting randomized experiment could be the golden standard, in which the experimental units are randomly assigned to treatment or control groups. As the treatment and control groups are identical on average except for the randomized assignment, we can interpret the differences in the outcome as the causal effect of the treatment.

Let D be the binary treatment variable. $D_i = 1$ indicates the experimental unit i is randomly assigned to the treatment group and $D_i = 0$ indicates the assignment to the control group. The corresponding potential outcome in the post-intervention period is defined as Y_{i1}^T and Y_{i1}^C, respectively (Y_{i0}^T and Y_{i0}^C in the pre-intervention period). Therefore, the treatment effect for i, τ_i, is the difference between

potential outcomes in the post-intervention period, which implies the pre-treatment outcomes are ignored. Formally, we have

$$\tau_i = Y_{i1}^T - Y_{i1}^C \tag{5}$$

Due to the so-called fundamental problem of causal inference (Holland 1986), it is not possible to evaluate the treatment effect for each signal experiment unit as we only can observe either the treated outcome Y_{i1}^T or controlled outcome Y_{i1}^C. However, in the context of randomized experiment, we can assess the average treatment effect (ATE) by means of comparing average outcomes across the treatment and control groups, which is referred as single difference estimator. That is,

$$ATE_D = E\left(Y_{i1}^T\right) - E\left(Y_{i1}^C\right) \tag{6}$$

It is worth noting that, in the context of non-randomized experiment, the true average treatment effect will be confounded by permanent differences between treatment and control groups that have existed prior to the intervention.

With regard to the non-randomized experiment, if both the pre-intervention and post-intervention data are available, the difference-in-difference estimator is most commonly applied in the program evaluation, which is defined as the difference in outcome in the treatment group pre- and post-intervention minus the difference in average outcome in the control group pre- and post-intervention. Formally, we have

$$ATE_{DD} = \left[E\left(Y_{i1}^T\right) - E\left(Y_{i0}^T\right)\right] - \left[E\left(Y_{i1}^C\right) - E\left(Y_{i0}^C\right)\right] \tag{7}$$

where the difference in the average outcome in the control group pre- and post-intervention $\left[E\left(Y_{i1}^C\right) - E\left(Y_{i0}^C\right)\right]$ captures the time trend. Therefore, an essential assumption of difference-in-difference method is that the treated and control groups have parallel time trend. Put differently, in absence of treatment, the average change of the outcomes would be the same for treated and control groups. Failure of this assumption would confound the estimated average treatment effect with a natural time trend producing biased inference. In practice, the closer the control and treatment groups in terms of both observable and unobservable characteristics, the greater the credibility of the difference-in-difference method in recovering the average treatment effect. We could check the parallel time trend assumption if more data on other time periods pre- and post-intervention are available.

Difference-in-difference method is simple but powerful to circumvent many endogeneity problems that typically result from making comparisons between heterogeneous experimental units (Meyer 1995). Nevertheless, we should be cautious to employ difference-in-difference method when many years of data are used and the outcomes are serially correlated since the standard errors are

inconsistent so as to result in serious over-estimation of t-statistics and significance levels (Bertrand et al. 2004).

The existence of heterogeneous treatment effect is well documented in the literature. The most common approach is to estimate separate average treatment effects across different subgroups (Bitler et al. 2010), which is termed as conditional average treatment effect (CATE) (Imbens & Wooldridge 2009). Let $\tau_i(x)$ denote the treatment effect of experimental units i conditional on having characteristics $X = x$. Therefore, the conditional average treatment effect is the difference between the conditional average outcomes across the treatment and control groups. Formally, for the single difference method and difference-in-difference method, respectively, we have

$$CATE_D = E(\tau_i(x)) = E\left(Y_{i1}^T \middle| X = x\right) - E\left(Y_{i1}^C \middle| X = x\right) \tag{8}$$

$$CATE_{DD} = E(\tau_i(x))) = \left[E\left(Y_{i1}^T \middle| X = x\right) - E\left(Y_{i0}^T \middle| X = x\right)\right] - \left[E\left(Y_{i1}^C \middle| X = x\right) - E\left(Y_{i0}^C \middle| X = x\right)\right] \tag{9}$$

5.2. Regression framework

Recall that the experimental unit of our intervention is household and two-wave pre- and post-intervention data are available. The benchmark difference-in-difference model is specified as follows.

$$Y_{it} = \alpha + \beta_0 X_{it} + \beta_1 T_i + \beta_2 t + \beta_3 (T_i \cdot t) + \varepsilon_{it} \tag{10}$$

where Y_{it} is the observed outcome for household i in the year t; $t = 0$ represents pre-intervention year and $t = 1$ represents the post-intervention year; X_{it} are the observable covariates; T_i identifies the treatment status of household i, which equals to 1 if in the treatment group and 0 if in the control group; β_0 contains the coefficients of covariates; β_1 captures the treatment group specific effect to account for the average permanent differences between the treated and control; β_2 represents the common time trend; β_3 indicates the true treatment effect; α is the constant term and ε_{it} is a random and unobserved error term.

The OLS estimators of regression (10) are unbiased only if the strict exogeneity assumption holds, i.e. all the regressors are uncorrelated with the error term ε_{it}. However, fixed-effect model has less strict assumption comparing with the simple OLS regression, where unobservable time-constant household-specific heterogeneity, γ_i, is allowed to freely correlated with regressors. Let $\varepsilon_{it} = \gamma_i + \eta_{it}$, so we have

$$Y_{it} = \alpha + \beta_0 X_{it} + \beta_1 T_i + \beta_2 t + \beta_3 (T_i \cdot t) + \gamma_i + \eta_{it} \tag{11}$$

where η_{it} is the error term of regression (11).

A within transformation is applied to get rid of the unobservable heterogeneity γ_i. By means of OLS, instead of estimating regression (10), fixed-effect model estimates the following regression

$$(Y_{it} - \overline{Y_i}) = \alpha + \beta_0(X_{it} - \overline{X_i}) + \beta_2(t - \bar{t}) + \beta_3(T_i \cdot t - T_i \cdot \bar{t}) + (\eta_{it} - \overline{\eta_i}) \quad (12)$$

where $\overline{Y_i}$, $\overline{X_i}$, $\bar{t}$ and $\overline{\eta_i}$ are the average of Y_{it}, X_{it}, t and η_{it}, respectively.

In the context of policy evaluation, if the treatment status is not randomly assigned to each household, the benchmark model has no chance to reveal the true treatment effect at all. In contrast, fixed-effect model allows limited endogeneity of treatment status. That is, the treatment status can be freely correlated with unobservable time-constant household-specific heterogeneity.

Specialized to our analysis, X_{it}, the observable covariates, are *han, low_edu, mid_edu, high_edu, hhsize children, wealth, ill*; T_i, the treatment status of household i, is *treat*. t, the year variable, is *post*.

6. Results

Table 7 summarizes the difference-in-difference results[64] of evaluating the impact of first-year provider payment intervention on out-of-pocket health expenditures. Three samples -unbalanced panel, balanced panel and sub-balanced panel- are employed in our analysis respectively. A set of variables related to out-of-pocket health expenditures are taken as outcome variables, which can be grouped into two categories: the absolute value of out-of-pocket health expenditures and the affordability of out-of-pocket health expenditure. We evaluate Haiyuan and Yanchi separately not only because they have significantly demographic varieties but also they did not share the same benefit package design of New Cooperative Medical Scheme (NCMS) previous to our intervention. We will firstly verify the robustness of the results with different combinations of samples and estimation methods; and then sum up the main results in terms of various outcome variables; at last, report the results of subgroup analysis to check whether the intervention affected the poor and non-poor heterogeneously.

6.1. The role of samples and estimation methods

The results using unbalanced panel is similar to that of using balanced panel. The average treatment effect tends to be larger in the latter but most of them are not significant. This is reasonable given the fact that there is no significant difference between the dropped-out and added-in households in terms of the demographic variables.

With regard to the sub-balanced panel, we are aware that excluding households with negative outpatient expenditure is debatable. However, there is a trade-off between sample bias and data

[64] Only the coefficients of treat*post under various settings are presented here. For the detailed results of the regressions, please see Table 8, Table 9 and Table 10 in the appendix.

reliability and validity. Many results obtained from evaluating sub-balanced panel are at least consistent with those from balanced panel, especially for the case of Yanchi. However, the affordability outcome variables of Haiyuan are the exceptions in that sense that the results from sub-balanced and balanced samples seem contradictory, although nothing significant in both cases. Furthermore, it is worth noting that the results from sub-balanced panel tend to be more significant or/and have larger quantitative values comparing with those from the balanced panel.

In terms of the estimation methods, i.e. OLS and fixed-effect, we find that, for any specific outcome variable and the given sample, the majority of the results are highly consistent except for the catastrophic health expenditure variable (*CHE_share*) of Haiyuan, which is actually not significant. The fact that OLS and fixed-effect estimation present the consistent results does make sense since our treatment status is randomly assigned to each town (in fact it is a cluster-randomized experiment), the OLS model still has chance to reveal the true treatment effect to some extent.

6.2. The results with regard to various outcome variables

Two measurements of outpatient expenditure –aggregating from two-week recall question as well as extracting from annual total health expenditure and inpatient expenditure- are adopted respectively in our analysis. It is surprising that, in the case of Haiyuan, these two measurements provide significantly contradictory results. While the two-week one shows that, for those who were followed up in Haiyuan, the two-week outpatient expenditure was significantly increased in the treatment towns thanks to the provider payment intervention; the one-year measurement shows the opposite. This contradiction exists regardless of the sample or the estimation method employed. However, Yanchi is not in this case in that the results from two measurements of outpatient expenditure are consistent, i.e. comparing with the control towns, the outpatient expenditure tends to be lowered in the treatment towns, although the significance and the size of the impact depend on the estimation method and the sample. In both of the two counties, total health expenditure tends to decrease, despite that most of them are not significant.

For the affordability-related outcome variables, the budget share of total health expenditure is significantly lowered and total expenditure net of health expenditure increases in Yanchi. It is understandable that there is almost no significant improvement in terms of 'catastrophic' health expenditure since the average treatment effects on budget share and total expenditure net of health expenditure are not necessarily beneficial to those who were below the threshold. However, there is nothing significant for Haiyuan and also the results swing with the choice of the sample.

Table 7: Impact of provider payment intervention

		unbalanced panel	balanced panel		sub-balanced panel	
		OLS	OLS	FE	OLS	FE
Haiyuan	*op_exp_2w*	45.1**	45.0*	45.0*	50.0**	49.6**
	op_exp_sub[65]	-140.0	-265.2**	-257.6**	-253.9**	-243.5**
	tot_exp	-86.7	-187.7	-174.8	-218.7	-203.3
	the_share	0.0013	0.00076	0.0021	-0.0032	-0.0015
	res_exp	80.8	30.0	23.6	-99.2	-114.5
	CHE_share	0.011	0.012	0.017	-0.0023	0.0032
	CHE_inc	-0.0047	-0.0024	-0.0019	0.0032	0.0038
	N	4996	3718	3718	3358	3358
Yanchi	*op_exp_2w*	-40.2	-47.9	-44.2*	-58.4*	-55.2*
	op_exp_sub	-78.1	-109.7	-90.6	-221.0**	-205.8**
	tot_exp	-132.8	-180.8	-147.8	-342.2**	-310.0**
	the_share	-0.017*	-0.021**	-0.017*	-0.028***	-0.023**
	res_exp	758.8**	770.7**	778.6**	846.2**	868.3**
	CHE_share	-0.048	-0.049	-0.035	-0.054	-0.038
	CHE_inc	-0.016	-0.016	-0.016	-0.027**	-0.026**
	N	2628	2260	2260	1962	1962

6.3. The heterogeneity of average treatment effect

Although it was not designed as pro-poor or pro-rich intervention, we are interested in checking whether this provider payment intervention generates different results across the population. The subgroup analysis is conducted using the balanced panel (Table 11) and sub-balanced panel (Table 12), respectively, to track the heterogeneity of average treatment effect. All the followed-up households are ranked according to their total expenditure in 2009. Those who ranked lower 50% are labelled as poor household. The others are labelled as non-poor households. We find that the intervention benefits the poor households more significantly in both counties. In fact, the significantly contradictory results with regard to the outpatient expenditure in Haiyuan seem to be less contradictory if we study the poor and non-poor households separately. For the non-poor households, there is a significant increase in terms of two-week outpatient expenditure but no significant decrease in terms of the one-year subtracted one. For the poor households, there is no significant increase in terms of two-week outpatient expenditure but there is significant decrease in terms of the one-year subtracted one.

[65] In this chapter, when *op_exp_sub* is taken as the outcome variable with unbalanced panel or balanced panel, all the negative *op_exp_sub* are replaced by the imputed values. This issue is not relevant to the sub-balanced panel since all the households with negative *op_exp_sub* have been dropped out.

Table 11: Impact of provider payment intervention (subgroup analysis with balanced panel)

		non-poor		poor	
		OLS	FE	OLS	FE
Haiyuan	*op_exp_2w*	79.6*	75.4*	12.1	11.6
	op_exp_sub	-205.5	-186.7	-353.0*	-360.4*
	tot_exp	-143.4	-99.2	-276.3	-280.8
	the_share	-0.0043	-0.0021	0.0024	0.0050
	res_exp	61.5	149.3	35.9	-44.4
	CHE_share	0.018	0.025	0.0047	0.011
	CHE_inc	-0.029	-0.034	0.023	0.027
	N	1860	1860	1858	1858
Yanchi	*op_exp_2w*	-7.56	-7.04	-93.5	-84.3*
	op_exp_sub	88.3	110.0	-330.4**	-303.0**
	tot_exp	-106.6	-78.7	-274.5	-219.7
	the_share	-0.00022	0.0015	-0.043***	-0.037***
	res_exp	86.0	140.9	1433.7***	1416.1***
	CHE_share	0.053	0.059	-0.15***	-0.13***
	CHE_inc	-0.016	-0.014	-0.014	-0.016
	N	1130	1130	1130	1130

Table 12: Impact of provider payment intervention (subgroup analysis with sub-balanced panel)

		non-poor		poor	
		OLS	FE	OLS	FE
Haiyuan	*op_exp_2w*	92.4*	87.3*	7.94	8.17
	op_exp_sub	-139.4	-119.3	-391.3*	-392.9*
	tot_exp	-139.6	-102.5	-332.5	-332.6
	the_share	-0.0088	-0.0057	-0.00041	0.0021
	res_exp	-179.4	-136.0	9.51	-71.4
	CHE_share	-0.0075	0.0029	-0.0013	0.0049
	CHE_inc	-0.025	-0.030	0.029	0.034
	N	1680	1680	1678	1678
Yanchi	*op_exp_2w*	-13.9	-13.8	-108.8	-97.7*
	op_exp_sub	-57.1	-43.7	-402.0**	-371.2**
	tot_exp	-302.3	-275.1	-415.8**	-369.5*
	the_share	-0.00076	0.0016	-0.056***	-0.049***
	res_exp	135.8	259.8	1589.0***	1555.9***
	CHE_share	0.068	0.077	-0.17***	-0.15***
	CHE_inc	-0.016	-0.015	-0.038**	-0.040**
	N	982	982	980	980

7. Concluding remarks

We attempt to evaluate the impacts of a provider payment intervention implemented in two counties – Haiyuan and Yanchi- in Ningxia Province (China). Difference-in-difference method is employed to estimate the average treatment effect with a set of outcome variables related to out-of-pocket health expenditure. The estimations were exercised repeatedly using unbalanced panel, balanced panel and sub-balanced panel, respectively, since each of them has some advantages and disadvantages in terms of the survey design and follow-up issues, data reliability and validity problems.

In general, the results of Yanchi are in line with the desired impacts of the intervention. Although some of them are not significant or the significance varies with the choice of the sample or estimation methods, they, at least, are self-consistent. Nevertheless, this is not the case of Haiyuan. As having been discussed, we find significantly contradictory results with regard to the outpatient expenditure and also all the results related to affordability are sensitive to the choice of the sample. However, the subgroup analysis suggests that the heterogeneity of treatment effect does exist in both counties. And on average, our analysis shows that the poor households benefited more from the provider payment intervention, although it was not designed as an ex-ante pro-poor intervention.

Recall that the design of the intervention is the same for two counties whereas the evaluation results diverse significantly. It may be caused by the following reasons. First, the two counties were distinct from each other in some ex-ante characteristic such as the economy and the pre-intervention health system. The descriptive statistics show that Haiyuan was much poorer than Yanchi in terms of the per-capita total expenditure in 2009 (less than 4000 RMB in Haiyuan while more than 7000 RMB in Yanchi) but it experienced a much higher growth rate in 2010 (around 25% in Haiyuan while 10% in Yanchi). Second, as having been mentioned in the previous section, what we have evaluated -the impact of provider payment intervention on the household health expenditure- is an incentive chain, the results of which depends on the sequential responses of providers and households, which may explain the instable results to some extent. Third, all of our findings are not conditional on the health utilization. Put differently, the findings mix up the change of health utilization and health expenditure conditional on utilization. The heterogeneity may be caused by different patterns in health utilization. Fourth, the occurrence of self-inconsistent results implies that there may be either some policy implementation issue or data quality problem in Haiyuan.

Furthermore, we should be cautious of interpreting the results since the one-year implementation period may be too short to show the clear or true impacts. It is recommendable to re-evaluate the impacts after a longer implementation period with more waves of panel data.

Table 1: Provider payment system

	Revenue	2009	2011	
Village clinic	Public health	2 RMB per person budget and fee for immunization /vaccination	adjusted capitation	= 7RMB * population in the village
	Basic primary care	fee for IV (zero drug profit)	adjusted capitation	= 13RMB per visit *1.5 visits per person * enrolled population in the village
			case-based	= 2 RMB * village clinic visits+ 4 RMB * home visits
	Government subsidy	1200 RMB	fix amount	1200 RMB
Township health centre	Public health	4 RMB per capita	adjusted capitation	= 8 RMB * population in the town
	Primary care	net revenue from IV, tests/exams, services (zero drug profit)	adjusted capitation	= expenditure cap per visit (differ across towns) * 1.5 visits per person * enrolled population in the town
	Inpatient care	fee-for-services	fee-for-service with cap	cap= inpatient revenue of 2009
	Government subsidy	salary for staff with an official slot[23]	fix amount	salary for staff with an official slot
County hospital	Health care	fee-for-services	fee-for-service with cap	cap for Haiyuan= (80 RMB*enrolled population-reimbursement to township health center in 2009)*41%; cap for Yanchi= min{(80 RMB*enrolled population-reimbursement to township health center in 2009)*55%, 4.988 million}

[23] Usually, only about 1/3 of the staffs have official slots. The rest are hired by contracts. Township health centres thus have to raise revenue to finance the contracted staff.

Table 2: Definitions of variables

Variable	Definition
post	Dummy variable =1 if in 2011
treat	Dummy variable =1 if in the intervention county of provider payment system
*post*treat*	Dummy variable =1 if in the intervention county of provider payment system in 2011
han	Dummy variable =1 if it is Han-ethnic household
low_edu	Dummy variable =1 if the highest education achievement within the household is primary school
mid_edu	Dummy variable =1 if the highest education achievement within the household is middle school
high_edu	Dummy variable =1 if the highest education achievement within the household is high school and above
ave_age	The average age of the adults with in the household
hhsize	The number of people living in the household
children	The number of children living in the household
wealth	Wealth index generated from household durable goods module
ill	Dummy variable =1 if at least one household member was ill in the past two weeks (self-assessed)
ip_need	Dummy variable =1 if at least one household member was suggested to be hospitalized in the past one year (diagnosed by the doctors)
food_exp	Per-capita household annual consumption on food including self-produced food (RMB)
the	Per-capita household annual total out-of-pocket health expenditure from single recall question in the household consumption module (RMB)
tot_exp	Per-capita household annual total expenditure constructed from household food expenditure, health expenditure and other expenditure (RMB)
ip_exp	Per-capita household annual out-of-pocket inpatient expenditure aggregated from individual level using recall questions in the health module (RMB)
op_exp_2w	Per-capita household two-week out-of-pocket outpatient expenditure (including self-purchased drugs) aggregated from individual level using recall questions in the health module (RMB)
op_exp_sub	Per-capita household annual outpatient expenditure (=*tot_exp-ip_exp*) (RMB)
the_share	Household total health expenditure share (=*the/tot_exp*)
res_exp	Per-capita household total expenditure net of health expenditure (=*tot_exp-the*) (RMB)
CHE_share	Dummy variable =1 if household total health expenditure accounts more than 10% of household total expenditure
CHE_inc	Dummy variable =1 if per-capita household total expenditure net of total health expenditure below local poverty line (1350 RMB)

Table 3: Principal components analysis of wealth index

Variable	Mean	Std. Dev.	Min	Max	Factor score
# of cows owned	0.296	1.125	0	35	0.014
# of sheep owned	11.173	27.607	0	400	0.079
# of horses, donkeys, mules owned	0.299	1.286	0	70	0.005
# of pigs owned	0.813	6.734	0	300	0.037
owns Xmu (1mu=.666hectares) irrigated lands	4.247	8.955	0	300	0.105
owns Xmu (1mu=.666hectares) fields on a mountain	16.832	19.876	0	471	-0.003
area of house in square meters	88.035	115.786	8	9016	0.076
owns watch/alarm clock	0.810	0.392	0	1	0.073
owns bicycle	0.301	0.459	0	1	0.080
owns radio	0.122	0.327	0	1	0.067
owns black & white tv	0.100	0.300	0	1	-0.004
owns color tv	0.901	0.298	0	1	0.114
owns sewing machine	0.588	0.492	0	1	0.122
owns motorbike	0.657	0.475	0	1	0.140
owns car	0.102	0.303	0	1	0.105
owns refrigerator	0.232	0.422	0	1	0.136
owns washing machine	0.588	0.492	0	1	0.155
owns telephone (including cell phone)	0.901	0.299	0	1	0.107
owns farming machine	0.505	0.500	0	1	0.110
owns VCR, DVD, sound system, TV receiver, etc.	0.618	0.486	0	1	0.110
owns camera, video camera, etc.	0.052	0.222	0	1	0.088
owns air conditioning	0.021	0.143	0	1	0.080
has a kitchen separated from living space	0.657	0.475	0	1	0.083
animals and people live in same room	0.047	0.211	0	1	0.019
Type of house					
concrete	0.281	0.450	0	1	-0.103
wood and brick	0.284	0.451	0	1	0.000
wood	0.262	0.440	0	1	-0.186
all brick	0.148	0.355	0	1	0.011
cave dwelling	0.020	0.141	0	1	-0.059

other	0.005	0.069	0	1	-0.030
Type of floor material					
dirt	0.121	0.326	0	1	-0.085
brick	0.647	0.478	0	1	0.000
tile	0.212	0.409	0	1	0.117
wooden planks	0.005	0.068	0	1	0.003
laminate	0.006	0.074	0	1	0.015
other	0.009	0.097	0	1	-0.005
Primary source of drinking water					
tap water	0.177	0.381	0	1	0.034
spring water	0.014	0.117	0	1	-0.021
cellar	0.573	0.495	0	1	-0.089
hand-drawn well or well water	0.186	0.389	0	1	0.000
river water	0.028	0.164	0	1	-0.012
ditch water	0.015	0.122	0	1	-0.014
other	0.007	0.085	0	1	-0.022
Type of toilet					
flush toilet	0.007	0.083	0	1	0.018
composting toilet that generates methane	0.006	0.079	0	1	0.016
a built structure	0.004	0.059	0	1	-0.002
pit	0.116	0.320	0	1	0.015
bucket	0.004	0.064	0	1	-0.003
hay	0.831	0.375	0	1	0.000
no toilet	0.027	0.162	0	1	-0.032
other	0.006	0.076	0	1	0.003
Source of cooking fuel					
coal	0.611	0.488	0	1	0.059
electric	0.170	0.375	0	1	0.054
kerosene	0.005	0.069	0	1	-0.003
natural gas	0.015	0.121	0	1	0.022
wood/hay/etc.	0.520	0.500	0	1	-0.074
charcoal	0.067	0.249	0	1	0.002
other	0.014	0.119	0	1	-0.009

Table 4: Probability of non-negative outpatient expenditure (*op_exp_sub*)

Variable name	Haiyuan	Yanchi
post	-0.00015	0.00014
hhsize	-0.014**	-0.038***
children	-0.0075	0.014
han	0.039***	0.092
low_edu	0.014	-0.036
mid_edu	0.0066	0.013
high_edu	0.011	0.000076
wealth	-0.00067	0.0098
ave_age	-0.0012**	-0.0024***
cons	1.00***	1.00***
N	3718	2260

* p<0.10 ** p<0.05 *** p<0.01

Table 5: Descriptive statistics (unbalanced/balanced sample)

Haiyuan	2009						2011					
	Full sample		Followed-up		Dropped-out		Full sample		Followed-up		Added-in	
	T	C	T	C	T	C	T	C	T	C	T	C
han	0.434	0.256	0.417	0.271	0.488	0.214	0.437	0.258	0.417	0.271	0.504	0.224
	(0.496)	(0.436)	(0.493)	(0.445)	(0.501)	(0.411)	(0.496)	(0.438)	(0.493)	(0.445)	(0.501)	(0.418)
low_edu	0.300	0.339	0.303	0.323	0.289	0.383	0.282	0.304	0.277	0.289	0.298	0.343
	(0.458)	(0.474)	(0.460)	(0.468)	(0.454)	(0.487)	(0.450)	(0.460)	(0.448)	(0.454)	(0.458)	(0.475)
mid_edu	0.374	0.396	0.379	0.407	0.360	0.368	0.370	0.404	0.365	0.403	0.384	0.406
	(0.484)	(0.489)	(0.485)	(0.492)	(0.481)	(0.483)	(0.483)	(0.491)	(0.482)	(0.491)	(0.487)	(0.492)
high_edu	0.266	0.197	0.260	0.205	0.285	0.176	0.286	0.225	0.299	0.240	0.240	0.184
	(0.442)	(0.398)	(0.439)	(0.404)	(0.452)	(0.382)	(0.452)	(0.418)	(0.458)	(0.427)	(0.428)	(0.388)
ave_age	33.71	31.69	34.28	32.01	31.82	30.82	35.00	33.10	35.17	34.00	34.44	30.72
	(16.98)	(15.30)	(17.31)	(14.68)	(15.71)	(16.82)	(16.49)	(15.16)	(16.17)	(15.36)	(17.55)	(14.37)
hhsize	4.415	4.560	4.414	4.611	4.417	4.426	4.278	4.443	4.320	4.472	4.136	4.368
	(1.524)	(1.471)	(1.512)	(1.450)	(1.566)	(1.518)	(1.503)	(1.575)	(1.499)	(1.593)	(1.511)	(1.526)
children	1.430	1.527	1.420	1.514	1.463	1.559	1.301	1.453	1.311	1.466	1.269	1.421
	(1.220)	(1.286)	(1.214)	(1.282)	(1.243)	(1.297)	(1.212)	(1.262)	(1.237)	(1.273)	(1.126)	(1.234)
ill	0.479	0.538	0.488	0.540	0.446	0.531	0.444	0.477	0.458	0.481	0.397	0.466
	(0.500)	(0.499)	(0.500)	(0.499)	(0.498)	(0.500)	(0.497)	(0.500)	(0.499)	(0.500)	(0.490)	(0.499)
ip_need	0.252	0.331	0.256	0.339	0.240	0.310	0.252	0.289	0.252	0.298	0.252	0.264
	(0.434)	(0.471)	(0.437)	(0.473)	(0.428)	(0.463)	(0.434)	(0.453)	(0.435)	(0.458)	(0.435)	(0.442)
tot_exp	3940.5	3674.3	3947.8	3707.7	3916.4	3585.8	5487.8	5256.0	5318.5	5270.0	6050.9	5218.9
	(4211.0)	(2762.0)	(4241.6)	(2808.4)	(4116.0)	(2636.1)	(5848.5)	(4270.2)	(4404.8)	(4059.1)	(9127.8)	(4791.2)
food_exp	1305.2	1235.2	1295.6	1232.8	1337.4	1241.6	1844.6	1754.0	1877.2	1774.8	1736.2	1698.7
	(930.0)	(779.5)	(913.5)	(777.8)	(984.3)	(785.0)	(1549.4)	(1388.4)	(1632.5)	(1329.8)	(1230.3)	(1533.6)
the	753.8	553.9	736.2	561.6	812.1	533.5	825.4	699.3	715.9	729.0	1189.9	620.4
	(3190.4)	(1146.5)	(3091.4)	(1153.8)	(3505.9)	(1128.4)	(3520.4)	(1761.8)	(1657.8)	(1942.1)	(6666.8)	(1152.1)
ip_exp	208.2	175.2	164.9	177.5	352.2	169.1	189.5	216.4	194.6	217.9	172.7	212.7
	(1670.4)	(689.8)	(832.5)	(688.7)	(3125.8)	(693.5)	(937.5)	(1279.2)	(979.8)	(1399.4)	(781.9)	(885.4)
op_exp_2w	52.77	86.72	49.88	79.90	62.41	104.8	51.00	37.39	57.86	40.30	28.18	29.68
	(210.1)	(418.4)	(164.0)	(380.8)	(319.0)	(505.0)	(481.8)	(179.8)	(546.0)	(202.4)	(110.6)	(96.80)
op_exp_sub	599.3	423.6	628.7	424.5	501.5	421.4	679.7	619.9	568.0	608.0	1051.0	651.3
	(2570.7)	(867.6)	(2859.0)	(846.9)	(1183.7)	(921.6)	(3245.6)	(2060.7)	(1182.0)	(1555.0)	(6393.7)	(3019.5)
the_share	0.134	0.132	0.132	0.132	0.140	0.130	0.123	0.117	0.119	0.117	0.135	0.114
	(0.174)	(0.161)	(0.172)	(0.161)	(0.181)	(0.160)	(0.165)	(0.149)	(0.160)	(0.151)	(0.181)	(0.144)
res_exp	3186.8	3120.4	3211.5	3146.0	3104.4	3052.3	4662.3	4556.7	4602.6	4541.0	4861.0	4598.5
	(2455.4)	(2429.8)	(2528.9)	(2466.0)	(2196.0)	(2332.5)	(4648.2)	(3837.1)	(3972.5)	(3543.5)	(6409.0)	(4529.8)
N	1047	1451	805	1054	242	397	1047	1451	805	1054	242	397

Yanchi	2009						2011					
	Full sample		Followed-up		Dropped-out		Full sample		Followed-up		Added-in	
	T	C	T	C	T	C	T	C	T	C	T	C
han	0.945	0.988	0.946	0.986	0.935	1	0.943	0.988	0.946	0.986	0.922	1
	(0.229)	(0.108)	(0.227)	(0.119)	(0.248)	(0)	(0.232)	(0.108)	(0.227)	(0.119)	(0.270)	(0)
low_edu	0.168	0.144	0.158	0.138	0.247	0.168	0.193	0.152	0.200	0.161	0.130	0.112
	(0.374)	(0.351)	(0.365)	(0.345)	(0.434)	(0.376)	(0.395)	(0.359)	(0.400)	(0.368)	(0.338)	(0.317)
mid_edu	0.499	0.493	0.501	0.493	0.481	0.495	0.449	0.466	0.433	0.466	0.584	0.467
	(0.500)	(0.500)	(0.500)	(0.500)	(0.503)	(0.502)	(0.498)	(0.499)	(0.496)	(0.499)	(0.496)	(0.501)
high_edu	0.291	0.324	0.298	0.328	0.234	0.308	0.291	0.321	0.301	0.320	0.208	0.327
	(0.454)	(0.469)	(0.458)	(0.470)	(0.426)	(0.464)	(0.454)	(0.467)	(0.459)	(0.467)	(0.408)	(0.471)
ave_age	42.66	39.37	42.25	40.01	46.12	36.45	40.62	39.51	40.53	39.93	41.41	37.62
	(19.29)	(17.02)	(18.63)	(17.48)	(24.02)	(14.47)	(15.52)	(15.53)	(15.08)	(15.46)	(18.86)	(15.79)
hhsize	3.641	3.916	3.654	3.926	3.532	3.869	3.706	3.853	3.730	3.901	3.506	3.636
	(1.462)	(1.438)	(1.446)	(1.467)	(1.594)	(1.304)	(1.490)	(1.453)	(1.485)	(1.468)	(1.527)	(1.369)
children	0.791	1.019	0.802	1.023	0.701	1	0.690	0.823	0.674	0.796	0.818	0.944
	(0.930)	(1.043)	(0.926)	(1.074)	(0.961)	(0.890)	(0.880)	(0.994)	(0.868)	(1.005)	(0.970)	(0.940)
ill	0.410	0.551	0.409	0.542	0.416	0.589	0.260	0.235	0.267	0.239	0.208	0.215
	(0.492)	(0.498)	(0.492)	(0.499)	(0.496)	(0.494)	(0.439)	(0.424)	(0.443)	(0.427)	(0.408)	(0.413)
ip_need	0.285	0.351	0.281	0.355	0.325	0.336	0.251	0.258	0.257	0.264	0.195	0.234
	(0.452)	(0.478)	(0.450)	(0.479)	(0.471)	(0.475)	(0.434)	(0.438)	(0.438)	(0.441)	(0.399)	(0.425)
tot_exp	7551.6	7013.8	7639.2	6975.5	6817.4	7187.0	8569.2	7397.0	8643.7	7421.9	7945.3	7284.6
	(5960.3)	(5716.6)	(5996.7)	(5684.1)	(5629.8)	(5885.4)	(5285.2)	(5077.2)	(5355.4)	(4748.6)	(4639.4)	(6383.5)
food_exp	3499.5	3403.7	3551.4	3435.0	3064.9	3262.0	3651.4	3117.6	3691.7	3160.3	3314.7	2924.0
	(3576.2)	(3989.0)	(3566.5)	(4100.8)	(3651.0)	(3450.9)	(3052.1)	(2671.1)	(3150.3)	(2485.9)	(2036.9)	(3393.7)
the	689.2	633.9	682.3	605.6	746.6	761.8	662.0	690.5	674.8	729.2	555.2	515.0
	(1676.1)	(1861.2)	(1706.3)	(1895.1)	(1406.4)	(1701.4)	(1187.1)	(1942.5)	(1216.5)	(2105.3)	(903.7)	(873.0)
ip_exp	174.9	217.3	167.5	212.4	237.0	239.6	227.6	494.4	240.9	431.5	116.1	779.5
	(821.5)	(1247.6)	(815.0)	(1329.6)	(876.6)	(778.4)	(887.9)	(3657.8)	(914.6)	(2506.0)	(613.6)	(6769.4)
op_exp_2w	37.56	43.84	38.44	44.31	30.15	41.68	36.86	61.73	39.96	70.56	10.86	21.69
	(145.6)	(182.7)	(151.9)	(197.4)	(73.93)	(89.71)	(231.1)	(510.7)	(243.5)	(562.6)	(58.76)	(83.68)
op_exp_sub	569.8	519.6	567.5	492.2	588.8	643.5	562.6	530.2	571.6	542.1	486.8	476.3
	(1382.5)	(1483.4)	(1428.1)	(1452.1)	(922.2)	(1619.2)	(985.2)	(1119.7)	(1006.9)	(1179.2)	(781.4)	(798.5)
the_share	0.0967	0.0923	0.0953	0.0880	0.109	0.112	0.0817	0.0875	0.0831	0.0894	0.0701	0.0792
	(0.136)	(0.139)	(0.135)	(0.130)	(0.142)	(0.176)	(0.120)	(0.136)	(0.122)	(0.140)	(0.0986)	(0.117)
res_exp	6862.4	6379.9	6956.9	6369.9	6070.8	6425.2	7907.1	6706.6	7968.9	6692.7	7390.1	6769.6
	(5568.0)	(5339.7)	(5623.2)	(5259.5)	(5045.6)	(5714.9)	(5136.0)	(4692.3)	(5204.2)	(4274.5)	(4521.9)	(6271.2)
N	722	592	645	485	77	107	722	592	645	485	77	107

mean coefficients; sd in parentheses
T: treated towns; C: control towns

Table 6: Descriptive statistics (negative/non-negative OP expenditure)

Haiyuan	2009						2011					
	Balanced panel		*op_exp_sub ≥0*		*op_exp_sub <0*		Balanced panel		*op_exp_sub ≥0*		*op_exp_sub <0*	
	T	C	T	C	T	C	T	C	T	C	T	C
han	0.417	0.271	0.425	0.280	0.338	0.196	0.417	0.271	0.425	0.280	0.338	0.196
	(0.493)	(0.445)	(0.495)	(0.449)	(0.477)	(0.399)	(0.493)	(0.445)	(0.495)	(0.449)	(0.477)	(0.399)
low_edu	0.303	0.323	0.309	0.317	0.235	0.366	0.277	0.289	0.284	0.288	0.206	0.304
	(0.460)	(0.468)	(0.463)	(0.466)	(0.427)	(0.484)	(0.448)	(0.454)	(0.451)	(0.453)	(0.407)	(0.462)
mid_edu	0.379	0.407	0.373	0.410	0.441	0.384	0.365	0.403	0.358	0.404	0.441	0.393
	(0.485)	(0.492)	(0.484)	(0.492)	(0.500)	(0.489)	(0.482)	(0.491)	(0.480)	(0.491)	(0.500)	(0.491)
high_edu	0.260	0.205	0.262	0.206	0.235	0.196	0.299	0.240	0.301	0.240	0.279	0.241
	(0.439)	(0.404)	(0.440)	(0.405)	(0.427)	(0.399)	(0.458)	(0.427)	(0.459)	(0.427)	(0.452)	(0.430)
ave_age	34.28	32.01	34.19	32.12	35.24	31.10	35.17	34.00	35.13	33.96	35.59	34.39
	(17.31)	(14.68)	(17.24)	(14.90)	(18.13)	(12.68)	(16.17)	(15.36)	(15.89)	(15.27)	(19.09)	(16.11)
hhsize	4.414	4.611	4.402	4.583	4.544	4.848	4.320	4.472	4.288	4.437	4.676	4.759
	(1.512)	(1.450)	(1.501)	(1.450)	(1.625)	(1.428)	(1.499)	(1.593)	(1.476)	(1.579)	(1.705)	(1.689)
children	1.420	1.514	1.399	1.499	1.647	1.643	1.311	1.466	1.281	1.458	1.632	1.536
	(1.214)	(1.282)	(1.191)	(1.288)	(1.422)	(1.229)	(1.237)	(1.273)	(1.219)	(1.278)	(1.392)	(1.237)
ill	0.488	0.540	0.478	0.540	0.603	0.536	0.458	0.481	0.455	0.477	0.500	0.518
	(0.500)	(0.499)	(0.500)	(0.499)	(0.493)	(0.501)	(0.499)	(0.500)	(0.498)	(0.500)	(0.504)	(0.502)
ip_need	0.256	0.339	0.214	0.291	0.706	0.741	0.252	0.298	0.223	0.252	0.574	0.688
	(0.437)	(0.473)	(0.411)	(0.454)	(0.459)	(0.440)	(0.435)	(0.458)	(0.416)	(0.434)	(0.498)	(0.466)
tot_exp	3947.8	3707.7	3986.2	3727.8	3531.4	3538.4	5318.5	5270.0	5254.3	5327.0	6013.6	4790.1
	(4241.6)	(2808.4)	(4347.9)	(2873.1)	(2836.2)	(2192.8)	(4404.8)	(4059.1)	(4257.2)	(4052.9)	(5762.6)	(4097.2)
food_exp	1295.6	1232.8	1298.6	1233.6	1262.3	1226.1	1877.2	1774.8	1857.8	1786.8	2087.0	1673.7
	(913.5)	(777.8)	(894.1)	(786.8)	(1109.5)	(700.2)	(1632.5)	(1329.8)	(1649.0)	(1352.1)	(1435.0)	(1125.0)
the	736.2	561.6	767.4	568.2	397.8	506.2	715.9	729.0	714.5	726.8	730.2	747.1
	(3091.4)	(1153.8)	(3224.3)	(1164.9)	(589.1)	(1058.3)	(1657.8)	(1942.1)	(1628.6)	(1785.9)	(1961.3)	(2957.0)
ip_exp	164.9	177.5	133.7	150.9	502.8	401.8	194.6	217.9	155.4	133.7	619.7	925.5
	(832.5)	(688.7)	(813.0)	(687.1)	(964.7)	(663.3)	(979.8)	(1399.4)	(848.1)	(615.5)	(1849.1)	(3847.3)
op_exp_2w	49.88	79.90	45.97	77.61	92.25	99.11	57.86	40.30	58.04	35.89	55.94	77.37
	(164.0)	(380.8)	(159.0)	(385.0)	(207.5)	(343.9)	(546.0)	(202.4)	(566.7)	(172.5)	(224.1)	(367.4)
op_exp_sub	628.7	424.5	633.7	417.4	574.2	484.5	568.0	608.0	559.2	593.1	663.8	733.3
	(2859.0)	(846.9)	(2977.7)	(825.2)	(826.7)	(1013.1)	(1182.0)	(1555.0)	(1111.0)	(1523.2)	(1787.8)	(1803.7)
the_share	0.132	0.132	0.134	0.133	0.104	0.128	0.119	0.117	0.120	0.120	0.108	0.0990
	(0.172)	(0.161)	(0.176)	(0.161)	(0.126)	(0.161)	(0.160)	(0.151)	(0.159)	(0.152)	(0.177)	(0.135)
res_exp	3211.5	3146.0	3218.7	3159.6	3133.5	3032.2	4602.6	4541.0	4539.8	4600.2	5283.4	4042.9
	(2528.9)	(2466.0)	(2523.5)	(2525.1)	(2604.6)	(1902.4)	(3972.5)	(3543.5)	(3787.5)	(3630.6)	(5592.5)	(2663.9)
N	805	1054	737	942	68	112	805	1054	737	942	68	112

Yanchi	2009						2011					
	Balanced panel		*op_exp_sub ≥0*		*op_exp_sub <0*		Balanced panel		*op_exp_sub ≥0*		*op_exp_sub <0*	
	T	C	T	C	T	C	T	C	T	C	T	C
han	0.946	0.986	0.952	0.988	0.904	0.970	0.946	0.986	0.952	0.988	0.904	0.970
	(0.227)	(0.119)	(0.214)	(0.109)	(0.297)	(0.173)	(0.227)	(0.119)	(0.214)	(0.109)	(0.297)	(0.173)
low_edu	0.158	0.138	0.151	0.138	0.205	0.136	0.200	0.161	0.189	0.158	0.277	0.182
	(0.365)	(0.345)	(0.359)	(0.346)	(0.406)	(0.346)	(0.400)	(0.368)	(0.392)	(0.365)	(0.450)	(0.389)
mid_edu	0.501	0.493	0.516	0.506	0.398	0.409	0.433	0.466	0.436	0.475	0.410	0.409
	(0.500)	(0.500)	(0.500)	(0.501)	(0.492)	(0.495)	(0.496)	(0.499)	(0.496)	(0.500)	(0.495)	(0.495)
high_edu	0.298	0.328	0.292	0.315	0.337	0.409	0.301	0.320	0.310	0.310	0.241	0.379
	(0.458)	(0.470)	(0.455)	(0.465)	(0.476)	(0.495)	(0.459)	(0.467)	(0.463)	(0.463)	(0.430)	(0.489)
ave_age	42.25	40.01	41.85	39.75	44.94	41.66	40.53	39.93	39.88	39.75	44.92	41.06
	(18.63)	(17.48)	(18.76)	(17.50)	(17.60)	(17.40)	(15.08)	(15.46)	(14.89)	(15.50)	(15.75)	(15.25)
hhsize	3.654	3.926	3.610	3.907	3.952	4.045	3.730	3.901	3.710	3.864	3.867	4.136
	(1.446)	(1.467)	(1.410)	(1.472)	(1.645)	(1.440)	(1.485)	(1.468)	(1.451)	(1.470)	(1.702)	(1.445)
children	0.802	1.023	0.817	1.026	0.699	1	0.674	0.796	0.687	0.792	0.590	0.818
	(0.926)	(1.074)	(0.925)	(1.090)	(0.934)	(0.977)	(0.868)	(1.005)	(0.874)	(1.013)	(0.827)	(0.959)
ill	0.409	0.542	0.391	0.532	0.530	0.606	0.267	0.239	0.260	0.227	0.313	0.318
	(0.492)	(0.499)	(0.489)	(0.500)	(0.502)	(0.492)	(0.443)	(0.427)	(0.439)	(0.419)	(0.467)	(0.469)
ip_need	0.281	0.355	0.249	0.301	0.494	0.697	0.257	0.264	0.183	0.189	0.759	0.742
	(0.450)	(0.479)	(0.433)	(0.459)	(0.503)	(0.463)	(0.438)	(0.441)	(0.387)	(0.392)	(0.430)	(0.441)
tot_exp	7639.2	6975.5	7748.0	6981.7	6903.1	6936.1	8643.7	7421.9	8716.0	7536.0	8154.2	6697.5
	(5996.7)	(5684.1)	(6095.1)	(5578.4)	(5255.8)	(6361.1)	(5355.4)	(4748.6)	(5278.3)	(4816.5)	(5862.0)	(4254.7)
food_exp	3551.4	3435.0	3602.9	3443.1	3202.7	3383.1	3691.7	3160.3	3767.1	3218.3	3180.9	2792.1
	(3566.5)	(4100.8)	(3632.1)	(4106.7)	(3082.8)	(4093.6)	(3150.3)	(2485.9)	(3143.9)	(2517.2)	(3164.6)	(2259.9)
the	682.3	605.6	718.4	528.5	438.0	1095.1	674.8	729.2	647.8	745.1	857.7	628.4
	(1706.3)	(1895.1)	(1803.7)	(1360.3)	(733.4)	(3816.0)	(1216.5)	(2105.3)	(1082.7)	(2215.1)	(1887.6)	(1199.0)
ip_exp	167.5	212.4	132.9	113.4	401.6	840.7	240.9	431.5	108.8	203.5	1135.1	1879.1
	(815.0)	(1329.6)	(742.3)	(876.7)	(1176.1)	(2785.1)	(914.6)	(2506.0)	(455.5)	(1446.5)	(2054.7)	(5553.6)
op_exp_2w	38.44	44.31	35.85	36.12	56.01	96.31	39.96	70.56	35.64	71.06	69.22	67.40
	(151.9)	(197.4)	(136.5)	(119.7)	(231.1)	(441.4)	(243.5)	(562.6)	(231.7)	(597.6)	(311.6)	(245.6)
op_exp_sub	567.5	492.2	585.5	415.1	445.7	982.0	571.6	542.1	539.0	541.6	792.5	545.7
	(1428.1)	(1452.1)	(1509.6)	(798.8)	(640.3)	(3364.0)	(1006.9)	(1179.2)	(907.4)	(1221.7)	(1507.2)	(869.5)
the_share	0.0953	0.0880	0.0980	0.0855	0.0771	0.103	0.0831	0.0894	0.0809	0.0874	0.0975	0.101
	(0.135)	(0.130)	(0.139)	(0.119)	(0.106)	(0.187)	(0.122)	(0.140)	(0.119)	(0.138)	(0.141)	(0.154)
res_exp	6956.9	6369.9	7029.5	6453.2	6465.0	5841.0	7968.9	6692.7	8068.2	6790.9	7296.5	6069.1
	(5623.2)	(5259.5)	(5689.3)	(5342.7)	(5158.7)	(4699.8)	(5204.2)	(4274.5)	(5159.6)	(4290.3)	(5482.3)	(4150.4)
N	645	485	562	419	83	66	645	485	562	419	83	66

mean coefficients; sd in parentheses
T: treated towns; C: control towns

Table 8: Regressions with unbalanced panel (OLS)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	-25.4**	173.1**	187.9**	0.0025	113.7	-0.025	0.0087
post	-45.9***	247.3***	174.3***	-0.0035	957.2***	-0.011	-0.050***
treat*post	45.1**	-140.0	-86.7	0.0013	80.8	0.011	-0.0047
hhsize	-16.0***	-142.5***	-169.2***	-0.0036*	-636.9***	0.0089	0.015***
children	2.89	52.1	34.8	-0.0031	-112.2**	-0.020***	0.0050
han	-27.6**	-38.0	11.4	0.0071	-669.8***	0.019	0.036***
low_edu	27.8**	-36.0	-90.0	-0.027**	216.8	-0.080**	-0.044**
mid_edu	47.3***	-19.7	11.3	-0.034***	655.0***	-0.12***	-0.073***
high_edu	49.2***	29.5	-14.9	-0.059***	1700.8***	-0.20***	-0.100***
wealth	2.96	-46.9	-72.5*	-0.023***	997.1***	-0.048***	-0.051***
ill	116.6***	303.5***	362.2***	0.047***	319.3***	0.16***	-0.030***
_cons	62.9***	838.2***	1076.3***	0.15***	5903.0***	0.40***	0.091***

NOTE: unbalanced sample: Haiyuan (N=4996)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	-5.54	194.5	271.5*	0.0045	11.9	-0.029	0.018
post	-20.2	298.7**	197.6**	-0.0015	1004.2***	-0.035	-0.053***
treat*post	53.5	-314.1*	-302.6	-0.0014	-134.0	0.028	-0.015
hhsize	-12.6**	-160.6***	-226.7***	-0.0061**	-597.3***	0.0026	0.019***
children	8.08	26.2	-18.8	-0.0019	-180.5**	-0.0091	0.0024
han	-34.8**	118.6	150.8	0.0089	-475.2***	0.028	0.038***
low_edu	-0.42	288.4	338.1	-0.011	331.6	-0.099**	-0.066**
mid_edu	27.7	129.1	334.9	-0.022	978.2***	-0.14***	-0.11***
high_edu	10.1	187.3	321.8	-0.039**	1651.9***	-0.18***	-0.12***
wealth	7.10	-63.0	-92.8	-0.022***	870.3***	-0.042***	-0.043***
ill	121.0***	258.8***	379.9***	0.059***	180.1	0.19***	-0.026**
_cons	46.2*	721.4***	1017.7***	0.14***	5728.6***	0.41***	0.095***

NOTE: unbalanced sample: Haiyuan & non-poor (N=2159)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	-45.0**	125.0	79.5	-0.0012	231.9*	-0.021	-0.0012
post	-65.3***	176.0***	179.6**	-0.0052	958.9***	0.015	-0.056***
treat*post	41.6*	-173.7	-140.9	-0.0025	121.3	-0.024	0.019
hhsize	-24.8**	-123.8***	-152.7***	-0.0030	-590.0***	0.016*	0.014***
children	1.33	12.5	28.9	-0.0037	-98.2*	-0.030***	0.010
han	-20.8	-78.8	-52.1	0.0075	-821.0***	0.00034	0.038**
low_edu	59.8***	-346.7**	-461.2**	-0.054***	238.5	-0.13**	-0.0032
mid_edu	77.3***	-289.8*	-351.4*	-0.056***	661.0**	-0.15***	-0.033
high_edu	95.8***	-249.5	-301.1	-0.083***	1794.9***	-0.24***	-0.069**
wealth	2.44	-6.27	-12.0	-0.022***	987.6***	-0.052***	-0.059***
ill	127.2***	239.2***	286.6***	0.038***	399.7***	0.14***	-0.033***
_cons	85.5**	1160.9***	1429.5***	0.18***	5541.9***	0.44***	0.060*

NOTE: unbalanced sample: Haiyuan & poor (N=2198)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	10.2	87.3	73.6	0.017**	-473.4*	0.070***	0.016*
post	57.0**	119.8*	179.2*	0.015**	-187.8	0.033	-0.0053
treat*post	-40.2	-78.1	-132.8	-0.017*	758.8**	-0.048	-0.016
hhsize	-9.24***	-94.1***	-141.7***	-0.0016	-1410.2***	-0.0043	0.0088***
children	2.49	-56.8*	-64.4*	-0.0081***	27.1	-0.032***	-0.0026
han	4.54	132.1*	153.8*	0.014	-115.8	0.045	0.0081
low_edu	-135.9	79.9	155.4	-0.028	1175.0***	-0.036	-0.039*
mid_edu	-126.7	13.5	61.0	-0.052***	2088.1***	-0.13***	-0.059***
high_edu	-129.6	-32.6	13.6	-0.073***	4021.8***	-0.20***	-0.062***
wealth	-1.14	-21.3	-22.1	-0.025***	1967.0***	-0.072***	-0.023***
ill	129.9***	345.6***	432.9***	0.050***	159.3	0.18***	-0.00047
_cons	125.6	618.5***	809.5***	0.12***	9164.7***	0.32***	0.045*

NOTE: unbalanced sample: Yanchi (N=2628)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	7.64	82.3	118.1	0.013	-800.3*	0.069**	0.012
post	100.8*	299.5***	406.5**	0.029**	-313.2	0.081**	-0.018
treat*post	-75.2	-176.3	-300.2	-0.019	988.7**	-0.072	-0.0071
hhsize	-16.2***	-63.5***	-113.2***	0.0016	-1514.1***	-0.0017	0.0090**
children	1.88	-107.1***	-116.6***	-0.013***	219.3	-0.033**	-0.0058
han	-16.8	63.1	126.7	-0.0096	394.1	0.0012	-0.016
low_edu	-284.0*	-3.68	218.9	-0.016	1011.0*	0.013	-0.0033
mid_edu	-261.9*	-138.4	-29.2	-0.049**	2073.0***	-0.12*	-0.023
high_edu	-261.2*	-332.5*	-224.2	-0.081***	4067.2***	-0.19***	-0.027
wealth	5.22	-6.68	-35.9	-0.027***	1916.9***	-0.063***	-0.028***
ill	177.1***	399.9***	455.9***	0.052***	488.5*	0.21***	0.0020
_cons	290.4*	718.9***	812.5**	0.13***	9028.4***	0.30***	0.045

NOTE: unbalanced sample: Yanchi & non-poor (N=1224)

	op_exp_2w	op_exp_sub	tot_exp	the_share	res_exp	CHE_share	CHE_inc
treat	11.2	82.8	18.9	0.020*	-109.0	0.068*	0.020*
post	26.6*	-50.7	15.4	0.0037	-3.87	-0.019	-0.00053
treat*post	-17.0	19.9	-18.9	-0.016	670.0	-0.025	-0.018
hhsize	-6.24*	-127.7***	-180.8***	-0.0045	-1385.9***	-0.0053	0.012***
children	6.21	-8.94	-12.1	-0.0034	-20.4	-0.027*	0.00065
han	11.1	155.6**	165.8*	0.028***	-204.4	0.065	0.031***
low_edu	9.01	239.9*	87.8	-0.047*	1455.6**	-0.091	-0.091**
mid_edu	14.2	240.5	175.1	-0.064**	2355.6***	-0.14*	-0.11***
high_edu	11.3	325.2**	256.0	-0.076***	4406.8***	-0.22***	-0.11***
wealth	-3.86	-30.1	7.31	-0.023***	2067.5***	-0.087***	-0.019***
ill	84.8***	272.4***	387.0***	0.046***	-149.7	0.15***	-0.0014
_cons	-18.3	515.7***	854.4***	0.13***	8759.5***	0.36***	0.052

NOTE: unbalanced sample: Yanchi & poor (N=1220)

Table 9: Regressions with balanced panel (OLS & FE)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	-22.8*		178.0**		147.9		-0.00095		86.5		-0.036		0.015	
post	-36.0***	-39.9**	229.2***	244.5***	164.4**	183.2**	-0.0052	-0.0088	951.9***	1082.2***	-0.018	-0.028	-0.052***	-0.058***
treat*post	45.0*	45.0*	-265.2**	-257.6**	-187.7	-174.8	0.00076	0.0021	30	23.6	0.012	0.017	-0.0024	-0.0019
hhsize	-12.3**	-14.4	-144.4***	-186.2***	-179.9***	-239.5***	-0.0032	-0.0045	-620.3***	-790.8***	0.0098	0.01	0.015***	0.022***
children	3.98	6.49	28.4	137.1*	15.8	200.1**	-0.0036	0.0072	-135.6***	-212.0*	-0.022***	0.029	0.0053	0.0056
han	-25.4**		25.3		14.1		0.0056		-640.0***		0.0084		0.030***	
low_edu	17.9	38.7	-84.6	-73.5	-169.5	-162.1	-0.046***	-0.013	448.7**	255.6	-0.13***	-0.023	-0.035	-0.046
mid_edu	39.3**	53.6	-146.7	-313	-155	-344.3	-0.052***	-0.024	959.1***	725.3*	-0.16***	-0.08	-0.077***	-0.059
high_edu	38.9**	75.6	-73.5	-228.7	-77.6	-322.8	-0.073***	-0.033	1953.8***	1123.0***	-0.22***	-0.098	-0.100***	-0.071*
wealth	2.29	14.1	-30.1	-68.7	-21.5	-54.7	-0.021***	-0.012**	949.6***	556.6***	-0.043***	-0.017	-0.043***	-0.028***
ill	114.4***	127.8***	239.7***	204.1**	301.1***	283.2***	0.045***	0.027***	282.2***	366.8***	0.16***	0.12***	-0.023***	-0.034***
_cons	46.5**	11	1004.8***	1220.6***	1327.5***	1519.1***	0.17***	0.14***	5613.2***	6467.7***	0.45***	0.30***	0.085***	0.068*

NOTE: balanced sample: Haiyuan (N=3178)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	-44.8*		2.7		-38.3		-0.018*		124.1		-0.087***		0.052**	
post	-45.5**	-49.1*	189.6*	195.7**	98.5	100.3	-0.0084	-0.011	981.6***	1045.8***	-0.0045	-0.013	-0.049***	-0.053***
treat*post	79.6*	75.4*	-205.5	-186.7	-143.4	-99.2	-0.0043	-0.0021	61.5	149.3	0.018	0.025	-0.029	-0.034
hhsize	-12.2	-8.41	-65.0*	-112.8*	-86.9*	-206.6***	0.0018	-0.0034	-727.2***	-950.4***	0.025**	0.012	0.022***	0.034***
children	4.88	6.34	5.58	69.6	-22.2	176.3	-0.0072*	0.0063	-145.8	-58.3	-0.024**	0.053*	0.012	-0.0017
han	-26.5		35.8		24		0.012		-916.1***		0.035		0.014	
low_edu	26.8	117	-354.3*	-396.3	-548.8**	-504.5	-0.064***	-0.017	605.9**	437.1	-0.17***	-0.0078	-0.032	-0.013
mid_edu	51.8*	106.3	-458.2**	-687.8**	-470.1**	-692.6**	-0.077***	-0.043	1335.8***	1299.3**	-0.23***	-0.12	-0.072**	0.00057
high_edu	53.3	146	-314.9	-619.1*	-352.3	-787.0**	-0.083***	-0.052*	2094.9***	1379.7*	-0.22***	-0.065	-0.091**	-0.0083
wealth	2.75	31.1	-26.7	-57.3	-5.67	-17.9	-0.022***	-0.014**	986.0***	606.7***	-0.043***	-0.0068	-0.059***	-0.044***
ill	139.2***	175.2***	318.7***	259.2**	360.0***	314.4***	0.047***	0.027***	354.5**	460.8**	0.17***	0.12***	-0.016	-0.027
_cons	38.9	-92.5	965.7***	1301.6***	1324.6***	1761.7***	0.18***	0.16***	5874.5***	6590.2***	0.44***	0.26***	0.049	-0.0067

NOTE: balanced sample: Haiyuan & non-poor (N=1860)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	-0.051		358.4**		347.1**		0.017		41.6		0.012		-0.015	
post	-23.4*	-24.4**	281.2***	315.1**	253.7**	293.7*	0.000078	-0.0065	913.1***	1100.6***	-0.029	-0.041	-0.056***	-0.066***
treat*post	12.1	11.6	-353.0*	-360.4*	-276.3	-280.8	0.0024	0.005	35.9	-44.4	0.0047	0.011	0.023	0.027
hhsize	-11.4***	-24.6***	-218.8***	-248.8***	-263.6***	-263.6***	-0.0069*	-0.0048	-532.4***	-660.2***	-0.0013	0.0059	0.011**	0.012
children	2.93	11.2	49.3	185.5	48.9	215.9	-0.0012	0.0072	-111.0**	-303.3**	-0.023*	0.01	-0.0017	0.01
han	-25.4***		22.2		15.3		0.00012		-345.3**		-0.018		0.040***	
low_edu	12.8	-29.2	126.9	180	136.4	121.5	-0.031	-0.0095	308.7	153.6	-0.086*	-0.031	-0.042	-0.083*
mid_edu	32.5*	10.1	122	-8.59	108	-48.2	-0.029	-0.0062	604.3**	266.7	-0.10**	-0.033	-0.084**	-0.12**
high_edu	29.8	16.7	144.3	78.7	164.5	88.4	-0.065***	-0.015	1861.4***	1053.5**	-0.23***	-0.14	-0.11***	-0.14**
wealth	-4.04	-7.13	-37.8	-94.8	-48.9	-109.5	-0.022***	-0.012	909.0***	516.0***	-0.049***	-0.026	-0.032***	-0.012
ill	89.6***	81.8***	164.9*	148	248.9**	249	0.043***	0.028***	223.1*	269.0*	0.16***	0.12***	-0.031**	-0.040**
_cons	40.7**	108.3***	1056.1***	1190.1**	1322.3***	1289.2**	0.16***	0.13***	5366.5***	6239.7***	0.45***	0.34***	0.12***	0.15***

NOTE: balanced sample: Haiyuan & poor (N=1858)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	10.8		113.5		90.4		0.019**		-346.7		0.067***		0.019**	
post	67.3**	61.0***	156.3*	131.6	239.7*	215.7*	0.020**	0.013	-100.3	273.9	0.037	0.012	-0.0056	-0.0064
treat*post	-47.9	-44.2*	-109.7	-90.6	-180.8	-147.8	-0.021**	-0.017*	770.7**	778.6**	-0.049	-0.035	-0.016	-0.016
hhsize	-11.0***	-18.4	-99.6***	-141.4***	-151.4***	-190.3***	-0.0017	-0.0034	-1437***	-1537.7***	-0.0036	0.00069	0.0090***	0.0053
children	4.04	-6.49	-54.0*	85.3	-59.5*	112.8	-0.0082***	0.00098	118.1	385.3	-0.031***	0.0023	-0.0011	-0.0033
han	0.81		139.6*		146.7		0.01		159.4		0.038		0.0042	
low_edu	-170.1*	-83	80.5	397.6*	120	597.9**	-0.040**	0.0042	1341.0***	315.5	-0.048	0.082	-0.057**	-0.038
mid_edu	-156.8	-96.0*	5.63	404.9*	30	789.6**	-0.063***	0.0077	2355.4***	986.4	-0.14***	0.099	-0.070***	-0.032
high_edu	-157.3	-105.3*	-54.3	369.9	-38.9	776.4**	-0.086***	-0.0013	4367.5***	1173.5	-0.22***	0.025	-0.072***	-0.023
wealth	-1.18	1.02	-29.3	28.1	-29.6	18.4	-0.024***	-0.013**	1924.9***	677.2***	-0.074***	-0.034*	-0.020***	-0.011*
ill	138.2***	127.6***	339.7***	185.9**	417.3***	202.3*	0.050***	0.024***	143.7	191.7	0.19***	0.099***	0.0015	0.0054
_cons	160.3*	149.2**	622.5***	528.9**	875.4***	442.2	0.13***	0.093***	8621.9***	10939.6***	0.33***	0.17**	0.053*	0.041

NOTE: balanced sample: Yanchi (N=2260)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	5.15		75.4		114.7		0.012		467.8		0.02		0.018	
post	21.7	21.8	-62	-87.4	134.9	112.4	-0.00073	-0.0082	322.2	599.4	-0.079*	-0.10***	0.009	0.0065
treat*post	-7.56	-7.04	88.3	110	-106.6	-78.7	-0.00022	0.0015	86	140.9	0.053	0.059	-0.016	-0.014
hhsize	-3.77	-1.3	-64.3	-136.6*	-118.9**	-194.1*	0.0017	0.0015	-1389***	-1323.8***	0.0082	0.013	0.0069*	0.0071
children	-1.94	7.85	-107.5**	107.8	-119.9**	147.1	-0.015***	0.0012	80.5	283	-0.060***	-0.0043	0.00022	-0.0044
han	-2.54		191.0**		208.1*		0.017		237.6		0.074		0.0042	
low_edu	-55.2	-4.67	232.6	483.1	457.6*	1084.2**	-0.025	0.023	1354.9**	322.1	-0.0058	0.12	-0.06	0.0081
mid_edu	-55.4	-20.5	120.6	526.2	300.6	1225.3***	-0.054**	0.01	2500.6***	653.9	-0.1	0.088	-0.084**	-0.021
high_edu	-49.1	-44.3	37.4	524	189	1254.6**	-0.082***	0.0069	4441.9***	180.2	-0.18**	0.0065	-0.083**	-0.019
wealth	-3.09	-3.23	-92.1**	-68.6	-116.8	-57.4	-0.029***	-0.017**	1705.2***	162.2	-0.082***	-0.044*	-0.023***	-0.020**
ill	105.9***	100.9***	215.6**	46.3	259.6**	67.8	0.043***	0.016	113.9	78.8	0.17***	0.076**	0.012	0.01
_cons	57.4	13.8	553.1***	534.5	658.8***	142.9	0.13***	0.085**	7784.2***	10792.0***	0.34***	0.21*	0.061	0.018

NOTE: balanced sample: Yanchi & non-poor (N=1130)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	17.8		184.5		108.3		0.030***		-1088***		0.12***		0.021	
post	110.3**	102.2**	365.4***	325.7***	316.2**	276.1*	0.040***	0.034***	-519.6	-126.9	0.15***	0.13***	-0.020*	-0.021
treat*post	-93.5	-84.3*	-330.4**	-303.0**	-274.5	-219.7	-0.043***	-0.037***	1433.7***	1416.1***	-0.15***	-0.13***	-0.014	-0.016
hhsize	-18.5***	-38.7*	-129.8***	-162.9**	-168.8***	-188.6**	-0.0041	-0.0095*	-1490***	-1708.3***	-0.013	-0.023	0.012***	0.0063
children	11.3	-24.3	1.65	67.1	-3.59	87.2	-0.0018	0.0013	190.6	416.4	-0.0066	0.015	-0.0015	0.00017
han	2.21		119.2		161.2		0.019		354.2		0.027		0.016**	
low_edu	-244.4	-138.2	-39.8	393.9	-194.2	170.3	-0.057**	-0.0069	1462.9***	398.2	-0.092	0.081	-0.066**	-0.088***
mid_edu	-219.9	-153.3	-63.8	356.1	-181	401.6	-0.070***	0.012	2290.6***	1262.9	-0.17***	0.15	-0.065**	-0.05
high_edu	-223.8	-139.8	-88	316.2	-197	357.7	-0.088***	-0.00008	4337.3***	2043.6*	-0.23***	0.086	-0.072**	-0.04
wealth	0.18	-0.82	13.7	100.7	34.1	77.6	-0.021***	-0.011	2103.2***	1148.2***	-0.075***	-0.04	-0.018***	-0.0013
ill	168.8***	147.2***	444.4***	287.5**	553.8***	298.6**	0.053***	0.029***	210.9	185.4	0.18***	0.10***	-0.0068	0.0051
_cons	231.6	280.9**	622.8***	558.2	942.5***	735.7*	0.12***	0.10***	9112.5***	11140.2***	0.29**	0.15	0.035	0.058

NOTE: balanced sample: Yanchi & poor (N=1130)

* p<0.10 ** p<0.05 *** p<0.01

Table 10: Regressions with sub-balanced panel (OLS & FE)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	-23.3		189.0*		171		0.0011		101.9		-0.028		0.011	
post	-38.1***	-42.7**	175.8***	187.0**	159.8**	173.4*	-0.0029	-0.0066	999.1***	1128.3***	-0.0058	-0.015	-0.053***	-0.061***
treat*post	50.0**	49.6**	-253.9**	-243.5**	-218.7	-203.3	-0.0032	-0.0015	-99.2	-114.5	-0.0023	0.0032	0.0032	0.0038
hhsize	-11.9**	-10.5	-162.2***	-205.5***	-198.2***	-261.9***	-0.0043	-0.0059	-619.6***	-759.8***	0.0089	0.012	0.013***	0.021***
children	4.37	1.51	28.5	168.2**	23.8	208.7**	-0.0029	0.0074	-145.4***	-273.0**	-0.019**	0.022	0.0055	0.0021
han	-18.4		37.2		28.3		0.0057		-639.3***		0.0097		0.032***	
low_edu	11.3	43.4	-16.3	12.2	-29.2	-7.33	-0.035**	-0.01	362.4*	302.9	-0.10***	-0.0025	-0.024	-0.048
mid_edu	30.7*	48.3	-74.5	-228.1	-58	-235.2	-0.043***	-0.024	866.7***	804.9**	-0.15***	-0.067	-0.064**	-0.061
high_edu	23.4	59.4	39.9	-98.5	72	-202.6	-0.061***	-0.029	1893.3***	1001.8**	-0.20***	-0.073	-0.093***	-0.076*
wealth	4.32	20.2	-31.9	-51.3	-35.6	-54.2	-0.022***	-0.015***	955.6***	590.4***	-0.048***	-0.028*	-0.043***	-0.023**
ill	110.3***	122.1***	222.7***	230.1***	328.3***	328.0***	0.047***	0.028***	307.0***	400.4***	0.17***	0.12***	-0.022**	-0.034***
_cons	51.6**	7.46	1004.3***	1163.8***	1262.3***	1484.1***	0.16***	0.15***	5681.9***	6383.9***	0.43***	0.29***	0.082***	0.084**

NOTE: sub-balanced sample: Haiyuan (N=3358)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	-46.2		1.81		-36.4		-0.017		156.7		-0.079**		0.048**	
post	-54.6**	-58.0*	118.1	121.2	90.9	92.9	-0.0059	-0.0086	1089.4***	1166.0***	0.0057	-0.0033	-0.053***	-0.059***
treat*post	92.4*	87.3*	-139.4	-119.3	-139.6	-102.5	-0.0088	-0.0057	-179.4	-136	-0.0075	0.0029	-0.025	-0.03
hhsize	-14.6	-9.61	-88.1**	-155.0***	-111.6**	-231.0***	0.00023	-0.0066	-723.3***	-857.1***	0.022**	0.0075	0.019***	0.033***
children	10.5	7.64	6.73	127.6	-6.17	182.7	-0.0068*	0.0062	-144.3	-208.3	-0.02	0.049	0.011	-0.017
han	-20.5		46.3		41.9		0.013		-919.8***		0.045		0.015	
low_edu	26.6	143	-352.8*	-383.4	-469.9**	-488.7	-0.053**	-0.018	489.5	467.6	-0.15**	0.0086	-0.02	-0.025
mid_edu	47.8	112.6	-391.0**	-643.6**	-438.2*	-709.1**	-0.067***	-0.045	1208.1***	1314.7**	-0.22***	-0.11	-0.055	-0.017
high_edu	43.9	136.3	-230.4	-529.5*	-259.2	-807.4**	-0.072***	-0.054*	1995.4***	916.6	-0.22***	-0.049	-0.085**	-0.032
wealth	4.7	37.4	-22.6	-26.8	-20.9	-15	-0.023***	-0.016**	971.8***	674.4***	-0.042***	-0.014	-0.058***	-0.034**
ill	139.0***	171.7***	279.7***	292.6***	402.9***	384.6***	0.048***	0.026***	379.1**	426.6*	0.17***	0.11***	-0.015	-0.028
_cons	45.9	-93.2	1027.3***	1343.3***	1332.8***	1849.0***	0.18***	0.18***	5934.0***	6508.9***	0.43***	0.28***	0.05	0.038

NOTE: sub-balanced sample: Haiyuan & non-poor (N=1680)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	0.86		377.8**		384.4**		0.019*		50.8		0.021		-0.021	
post	-17.7*	-20.2*	243.0**	272.2*	244.7**	276.2*	0.0019	-0.0045	899.3***	1083.9***	-0.012	-0.023	-0.055***	-0.065***
treat*post	7.94	8.17	-391.3*	-392.9*	-332.5	-332.6	-0.00041	0.0021	9.51	-71.4	-0.0013	0.0049	0.029	0.034
hhsize	-8.78**	-16.0**	-231.3***	-251.2***	-275.5***	-284.9***	-0.0073*	-0.0042	-537.7***	-674.3***	-0.00042	0.014	0.0100*	0.011
children	-1.1	2.31	47.6	198	48.3	221.8	-0.0004	0.0074	-129.4**	-324.1**	-0.021*	0.002	-0.000077	0.016
han	-17.8**		38.7		27.6		-0.00039		-332.8**		-0.022		0.041**	
low_edu	3.1	-40.8	252.3	374.9	323.9	434.4	-0.021	-0.0029	235.3	161.1	-0.063	-0.0067	-0.03	-0.077
mid_edu	21.9	-5.61	189.3	152.8	255	201.9	-0.022	-0.0054	529.5*	363.8	-0.081	-0.024	-0.075**	-0.11**
high_edu	10.2	-4.11	277.8	291.7	357.1	380.6	-0.054***	-0.0059	1835.9***	1263.1**	-0.20***	-0.11	-0.10***	-0.13**
wealth	-2.44	-1.62	-43.1	-89.9	-60.8	-106.7	-0.025***	-0.014*	933.5***	532.1***	-0.062***	-0.042*	-0.032***	-0.012
ill	82.0***	75.0***	165.2*	159.6	253.6**	261	0.045***	0.030***	252.6**	358.6**	0.17***	0.13***	-0.030**	-0.040**
_cons	41.8**	96.7***	1006.8***	1012.9**	1207.0***	1133.4**	0.15***	0.12***	5466.3***	6203.0***	0.41***	0.29***	0.11***	0.14**

NOTE: sub-balanced sample: Haiyuan & poor (N=1678)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	18.9**		200.9***		216.5**		0.025***		-412.7		0.078***		0.031***	
post	75.5**	71.4***	228.7***	214.9***	332.5***	320.8***	0.020**	0.013	-136.8	190	0.032	0.0058	-0.00021	-0.002
treat*post	-58.4*	-55.2*	-221.0**	-205.8**	-342.2**	-310.0**	-0.028***	-0.023**	846.2**	868.3**	-0.054	-0.038	-0.027**	-0.026**
hhsize	-8.55***	-19.5	-81.0***	-96.6**	-122.1***	-144.6**	0.00059	0.00085	-1480.2***	-1610.3***	0.0012	0.012	0.0080***	0.0061
children	6.1	-3.82	-59.1*	74.1	-60	108.6	-0.01***	-0.0025	190.7	520.8*	-0.033***	-0.0047	-0.0045	-0.013
han	2.79		148.6*		176.1*		0.013		-99.1		0.046		-0.0047	
low_edu	-154.8	-96.1*	151.1	478.4**	146.7	766.4***	-0.048**	0.011	1432.7***	141.4	-0.049	0.13*	-0.071**	-0.040*
mid_edu	-138.3	-100.4	25	343	-32.6	785.7**	-0.075***	0.012	2403.0***	646.5	-0.15***	0.14*	-0.084***	-0.036
high_edu	-143.4	-114.8*	-59	231.5	-79	773.4**	-0.096***	0.0021	4427.1***	1136.1	-0.22***	0.077	-0.086***	-0.036
wealth	-4.99	-3.01	-44.7	-20.6	-40.9	-31.4	-0.024***	-0.015***	1907.7***	787.2***	-0.072***	-0.036*	-0.018***	-0.01
ill	131.7***	128.8***	353.7***	214.8***	414.4***	201.4*	0.051***	0.024***	66.9	4.75	0.20***	0.097***	-0.0019	0.0024
_cons	127.7	155.3**	448.6***	401.4*	699.6***	251.9	0.13***	0.077***	9014.0***	11362.2***	0.31***	0.1	0.073**	0.052*

NOTE: sub-balanced sample: Yanchi (N=1962)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	13.1		189.3*		252.7**		0.013		480.6		0.018		0.02	
post	19.2	20.4	23.5	2.01	239.2	199.4	-0.0053	-0.015	268.3	474	-0.11**	-0.14***	0.0064	0.0035
treat*post	-13.9	-13.8	-57.1	-43.7	-302.3	-275.1	-0.00076	0.0016	135.8	259.8	0.068	0.077	-0.016	-0.015
hhsize	0.29	-0.33	-52.9	-65.6	-76.5	-112.3	0.0054	0.0068	-1517.2***	-1371.3***	0.016	0.025	0.0033	0.0077
children	-1.2	6.65	-126.8**	70.3	-139.4**	114.3	-0.019***	-0.0059	189.5	411.1	-0.069***	-0.014	-0.00062	-0.016
han	-7.39		195.6**		247.8*		0.02		179.2		0.087		-0.01	
low_edu	1.88	-11.6	317.4*	667.3**	415.4	1411.7***	-0.053	0.032	1765.8**	160.4	-0.028	0.17*	-0.084	0.0079
mid_edu	3.03	-15.4	163.1	502.7*	159.6	1377.9***	-0.085***	0.024	3063.4***	432.1	-0.13	0.15	-0.11**	-0.0091
high_edu	4.3	-59.5	74.4	401.8	55.2	1340.2**	-0.11***	0.015	5028.6***	-176.5	-0.22**	0.071	-0.11**	-0.022
wealth	-1.6	6.04	-84.4**	-78.9	-126.7	-73.8	-0.028***	-0.015*	1683.3***	210.1	-0.081***	-0.041	-0.020***	-0.017**
ill	86.4***	90.1***	183.4**	16.3	162.2	-90.7	0.037***	0.0038	246.8	-9.91	0.16***	0.04	0.0098	0.0077
_cons	-9.8	8.85	386.4**	291.9	547.9**	-234.6	0.15***	0.067*	7785.8***	11255.5***	0.36***	0.15	0.11**	0.02

NOTE: sub-balanced sample: Yanchi & non-poor (N=982)

	op_exp_2w		op_exp_sub		tot_exp		the_share		res_exp		CHE_share		CHE_inc	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
treat	24.8		236.0**		223.4		0.040***		-1194.0***		0.14***		0.042***	
post	124.3**	121.3***	401.2***	379.7***	380.7**	372.0**	0.043***	0.039***	-520.7	-193.9	0.16***	0.14***	-0.0065	-0.0084
treat*post	-108.8	-97.7*	-402.0**	-371.2**	-415.8**	-369.5*	-0.056***	-0.049***	1589.0***	1555.9***	-0.17***	-0.15***	-0.038**	-0.040**
hhsize	-16.9***	-42.8*	-98.4***	-127.9*	-144.6***	-152.1*	-0.0027	-0.0056	-1452.9***	-1860.2***	-0.0088	-0.013	0.013***	0.0065
children	13.5*	-20.9	5.6	66.7	8.22	81.3	-0.0022	0.000052	210.3	512.7	-0.0053	0.0079	-0.0063	-0.007
han	11.5		79.7		78.2		0.023		-844.3		0.039		0.019*	
low_edu	-239.5	-147.2	16.3	377.2	-123.8	222.1	-0.054**	-0.00051	1361.9**	354.3	-0.083	0.13	-0.078**	-0.095***
mid_edu	-213.7	-154.8	-70.6	241.8	-200.3	263.4	-0.071***	0.011	1922.1***	821.3	-0.17***	0.17*	-0.073**	-0.069*
high_edu	-220.8	-133.3	-129.9	145.4	-167.3	292.4	-0.084***	0.0013	3966.7***	2287.3*	-0.22***	0.14	-0.081**	-0.062
wealth	-7.93	-19.7	-14.2	26.3	33.8	5.89	-0.021***	-0.018**	2093.5***	1428.5***	-0.071***	-0.049*	-0.016***	-0.0012
ill	174.3***	154.3***	508.5***	365.3***	656.5***	447.1***	0.061***	0.038***	-88.6	-151.3	0.21***	0.12***	-0.01	0.002
_cons	206.3	294.7**	488.9**	510.3	814.5**	646.3	0.100***	0.086***	10495.0***	11685.8***	0.24	0.073	0.028	0.079**

NOTE: sub-balanced sample: Yanchi & poor (N=980)

* p<0.10 ** p<0.05 *** p<0.01

Reference

Ahmed, N., Brzozowski, M. & Crossley, T.F. (2006). Measurement errors in recall food consumption data, *Institute for Fiscal Studies Working Paper W06/ 21*.

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

Bitler, M., Gelbach, J. & Hoynes, H. (2010). Can constant treatment effects within subgroup explain heterogeneity in welfare reform effects? Available at: www.econ.ucdavis.edu/faculty/hoynes/working_papers/bgh-subgroups-paper.pdf

Blisard, N., & Blaylock, J. (1993). Distinguishing between market participation and infrequency of purchase models of butter demand. *American Journal of Agricultural Economics*, *75*(2), 314-320.

Feldstein, M. (1973). The welfare loss of excess health insurance. *Journal of Political Economy*, *81*(2), 251-280.

Filmer, D. & Pritchett, L. (2001). Estimating wealth effects without expenditure data -- or tears: An application to educational enrolments in states of India. *Demography, 38,* 115-132.

Holland, W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81(396),* 945-960.

Hsiao, W., Yip, W., & Wang, H. (2008). A social experiment in rural China: Rural Mutual Health Care–a summary report. *Harvard Program in Health Care Financing Working Paper*.

Hu, S., Tang, S., Liu, Y., Zhao, Y., Escobar, M.-L., & de Ferranti, D. (2008). Reform of how health care is paid for in China: challenges and opportunities. *The Lancet*, *372*(9652), 1846-1853.

Imbens, G. & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47(1),* 5-86.

Lu, C., Chin, B., Li, G., & Murray, C. J. L. (2009). Limitations of methods for measuring out-of-pocket and catastrophic private health expenditures. *Bulletin of the World Health Organization, 87(3),* 238-244.

Meyer, B. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics, American Statistical Association, 13(2),* 151-61.

McGuire, T. (2000). Physician agency. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics*. Amsterdam: Elsevier.

McGuire, T. G., & Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics*, *10*(4), 385-410.

Miller, H. D. (2009). From volume to value: better ways to pay for health care. *Health Affairs*, *28*(5), 1418-1428.

Thalmann, P. (2003). “House poor” or simply “poor”? *Journal of Housing Economics*, *12*, 291-317.

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.

Sanchis-llopis, J. A. (2001). Infrequency of purchase: A model for clothing consumption with panel data. *Annales d’Economie et de Statistique*, (61), 5.

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

Wagstaff, A., & van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. *Health Economics*, *12*(11), 921-34.

Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J. L. (2003). Household catastrophic health expenditure: a multicountry analysis. *The Lancet*, *362*(9378), 111-117.

Yip, W. C.-M., Hsiao, W., Meng, Q., Chen, W., & Sun, X. (2010). Realignment of incentives for health-care providers in China. *The Lancet*, *375*, 1120-1130.

Yip, W., & Hsiao, W. (2009). China's health care reform : a tentative assessment. *China Economic Review*, *20*(4), 613-619.