Available online at www.sciencedirect.com
ScienceDirect
Journal of Health Economics 27 (2008) 990–1005
JOURNAL OF HEALTH ECONOMICS
www.elsevier.com/locate/econbase

Can insurance increase financial risk? The curious case of health insurance in China

Adam Wagstaff*, Magnus Lindelow

The World Bank, 1818 H Street NW, Washington, DC 20433, USA

Received 17 October 2005; received in revised form 13 December 2007; accepted 1 February 2008
Available online 9 February 2008

Abstract

We analyze the effect of insurance on the probability of an individual incurring 'high' annual health expenses using data from three household surveys. All come from China, a country where providers are paid fee-for-service according to a schedule that encourages the overprovision of high-tech care and who are only lightly regulated. We define annual spending as 'high' if it exceeds a threshold of local average income and as 'catastrophic' if it exceeds a threshold of the household's own per capita income. Our estimates allow for different thresholds and for the possible endogeneity of health insurance (we use instrumental variables and fixed effects). Our main results suggest that in all three surveys health insurance increases the risk of high and catastrophic spending. Further analysis suggests that this is due to insurance encouraging people to seek care when sick and to seek care from higher-level providers.

JEL classification: G22; I1

Keywords: Health insurance; Financial risk; China

1. Introduction

The most basic argument for insurance is that it reduces financial risk. The classic textbook argument in the case of health insurance has an individual facing a known probability of falling ill and a corresponding known reduction in wealth caused by the medical expenses necessitated by falling ill (cf. e.g. Culyer, 1989). If offered full and actuarially fair insurance the risk-averse individual accepts it, preferring to pay the corresponding premium thereby securing a certain wealth equal to the expected wealth in the absence of insurance. The benefit of insurance is the reduction in risk—the knowledge that whether or not illness occurs, wealth is the same in both states. Relaxing the assumption that insurance offered is full leaves the risk-averse individual preferring insurance, because although not eliminated, the risk associated with illness is substantially reduced.

How this characterization of health insurance plays out in practice – and therefore how far health insurance protects people from financial risk – has been the subject of very little empirical research. Yet it is not obvious that in the real world health insurance always reduces risk. Contrary to the textbook example, there is not a fixed financial loss associated with illness, or even with each type of illness. A wide variety of tests and interventions can be undertaken,

* Corresponding author. Tel.: +1 202 473 0566; fax: +1 202 522 1153.
E-mail address: awagstaff@worldbank.org (A. Wagstaff).

0167-6296/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhealeco.2008.02.002

even for patients with similar conditions. Patients are not indifferent to the type and extent of care they receive, because in contrast to the textbook model, they derive utility from health status as well as financial wealth, and additional tests and interventions may be expected – at least up to a point – to increase the chances of a recovery. So, patients have an incentive to engage in ex post-moral hazard, increasing their demand for care as the price is reduced through insurance (Feldstein, 1973). This would not be sufficient to raise out-of-pocket payments, of course, unless demand were price elastic, which empirical work from around the world suggests it is not. If, however, insurance were to cause providers to shift the patient's demand curve to the right, out-of-pocket payments could increase as a result of having insurance. Inasmuch as the demand curve indicates the amount of care the patient would choose at a given price if he had the same information as the provider, and given that providers gain by such action, such a shift amounts to demand inducement (McGuire, 2000). In the McGuire-Pauly model (McGuire and Pauly, 1991) insurance might be argued to reduce the disutility a provider experiences from a given amount of inducement on the grounds that health care becomes less burdensome financially for patients with insurance; the effect would be to increase the provider's optimal level of inducement. This outcome seems less likely the greater is the control that the insurer has over the care delivered by the provider for a given medical condition (e.g. by imposing quality standards), the less freedom providers have to set prices, the greater the degree of self-regulation by the medical profession, and the stronger is any ombudsman or other authority acting on behalf of patients.

In many countries, especially developing countries, these checks on provider behavior are typically very limited if not largely absent. This is largely true of China, which is the setting for the present paper. Systems for monitoring and enforcing quality standards are weak. Providers are restricted in the prices they can charge patients. However, the government-set schedules for fees and medicines provide physicians with a strong incentive to favor high-tech care over basic care. For basic interventions, the government has set the price below cost so as to make them affordable even to fairly poor patients, while more sophisticated interventions are priced above cost to enable providers to make profits on them in the hope that providers will use these profits to cross-subsidize the delivery of basic interventions. In practice, and contrary to the outcome hoped for by the government, the price structure encourages providers to supply sophisticated care wherever possible, by shifting demand from low-margin basic services to high-margin high-tech care and drugs. Unsurprisingly, even low-level facilities have acquired sophisticated medical equipment, and there is evidence the care the system delivers is more costly and more sophisticated than is medically necessary (cf. Liu and Mills, 1999). The incentive to over-treat is accentuated where, as in China, there is a third party picking up part of the cost, especially one that is simply reimbursing (a fraction of) the costs incurred by the provider. Self-regulation by the Chinese medical profession is limited, and while the Chinese government has identified the delivery of unnecessary and poor quality care as a matter for concern, there are no formal complaint procedures for patients who feel that they have been over-treated. In such a setting, it seems perfectly plausible that at least some patients may end up not only getting more care than would have been the case if they had been uninsured, but also paying more out of pocket. Insurance in such a setting may, in other words, actually *increase* the probability of large out-of-pocket payments and hence exposure to financial risk.

Some studies to date have, of course, looked at the effect of insurance on out-of-pocket spending. Often these are tabulations of spending by insurance status or cross-section regressions of spending on insurance status and other covariates: examples from the US, which find *higher* out-of-pocket spending among the insured, include Rubin and Koelln (1993), Waters et al. (2004) and Shen and McFeeters (2006).[1] Such studies are vulnerable to biases caused by health insurance being endogenous. Sepehri et al. (2006) find that in Vietnam failure to take into account endogeneity (or equivalently selection on unobservables) causes the impact on out-of-pocket spending to be biased upwards, in their case making the difference between insurance having no impact on out-of-pocket spending and having a significant dampening effect. Finkelstein and McKnight (2005) use differences-in-differences and find that Medicare in the US significantly reduced out-of-pocket spending. A different approach is taken by Gertler and Solon (2000), who examine hospital pricing behavior vis-à-vis insured and uninsured patients in the Philippines. They find that private hospitals earn larger revenues from insured patients: for insured patients, hospitals secure payments from the insurer, but manage to earn similar out-of-pocket revenues to those they earn from uninsured patients. Gertler and Solon conclude that this is due to hospitals charging higher prices for insured patients, not delivering more services.

[1] The volume edited by Preker and Carrin (2003) contains several studies in this genre from developing countries.

We estimate the impact of insurance on the probability of households in China incurring 'large' out-of-pocket payments. We define 'large' in relation to the household's own per capita income (or consumption) and in relation to the local average income (or consumption). We also estimate the impact of insurance on the actual value of out-of-pocket payments, though the binary variable approach arguably gets better at the issue of how insurance protects against the risk of large out-of-pocket payments.[2] We use data from three separate household surveys spanning the period 1991–2004 and covering half of China's 32 provinces. We pay particular attention to the problem of endogeneity.

2. Health insurance in China—some salient facts

Under China's pre-reform planned economy, almost all citizens were covered by some form of health insurance. Agricultural workers were covered by the old commune-based cooperative medical scheme (CMS), state-owned enterprise (SOE) workers were covered by the Labor Insurance Scheme (LIS), and civil servants and other government workers were covered by the Government Insurance Scheme (GIS). There were some gaps in coverage (not all urban schemes covered dependents, for example), but the gaps were relatively small (during the 1970s the CMS covered an estimated 90% of the rural population).

China's transition from a planned to a market economy from 1980 onwards brought dramatic reductions in health insurance coverage. The decollectivization of agriculture resulted in an almost total collapse of the CMS. By 1993 less than 10% of the rural population had health insurance.[3] The mid-late 1990s saw several attempts to resuscitate the CMS, but despite these initiatives, CMS coverage nationally remained stubbornly low. By 2003 80% of China's rural population – some 640 million people – lacked health insurance. In that year, half of the rural residents in 2003 who said they did have insurance said they were covered by either private (i.e. commercial) insurance or 'other' insurance, up from 31% in 1993.

Coverage in China's cities has also declined, though less dramatically than in rural areas. As China transitioned to a market economy, the SOE – the backbone of the LIS – came under increasing pressure. LIS coverage fell, as did GIS coverage. By 1998 nearly half the urban population lacked insurance coverage. A variety of reforms have been introduced, including the setting up in 1998 of a new single urban scheme known as Basic Medical Insurance (BMI), into which LIS and GIS are gradually being subsumed. Despite these reforms, coverage by GIS/LIS/BMI continued to fall between 1998 and 2003. Had it not been for the growth of private and 'other' insurance schemes – these together covered 30% of the urban insured in 2003, compared to just 10% in 1993 – coverage in urban China would have fallen below 40% in 2003.

Until fairly recently, providers in China – even those delivering care to insured patients – were paid on a FFS basis. Those CMS schemes that survived during the 1990s simply reimbursed their members' medical bills, and made little attempt to restrict their choice of provider. Prior to the reform process that started in the mid-1990s, the same was true of the LIS and GIS. Members paid very little out of their own pockets, and providers were paid on a FFS basis. Unsurprisingly, costs increased rapidly, exacerbated by the continuous introduction of ever-more-costly medical technology. In 1995, the government launched a health insurance pilot experiment in the cities of Zhenjiang (Jiangsu province) and Jiujiang (Jiangxi province) (cf. e.g. Liu et al., 2001, 2003). Among the key elements of the reform was the setting-up of a citywide insurance pool across all work units, financed jointly by employers and employees. The pooled funds were then distributed into individual medical savings accounts (MSA) and a social pooling account (SPA). Supply-side cost-sharing was also introduced, although not until 1997 in Zhenjiang. At the end of 1996, the experiment reform was extended to 57 other cities, and in December 1998 the government called for reform of the existing GIS and LIS in China's remaining cities. By the end of 2003, the vast majority of large cities had implemented

[2] Waters et al. (2004), Shen and McFeeters (2006) and O'Donnell et al. (2005) look at the impacts of insurance on a binary variable capturing especially large spending. None, however, captures the endogeneity of insurance. Finkelstein and McKnight (2005) look at the impact on the mean and variance of the out-of-pocket payment distribution, using differences-in-differences. They use their results to estimate the risk-reduction benefits associated with Medicare.

[3] Figures on coverage are from China's National Health Survey (NHS), undertaken by the Ministry of Health (MOH) in 1993, 1998 and 2003. The figures for 1993 and 1998 are published in Gao et al. (2002). The urban insurance figures for 2003 are taken from "Health Services Utilization and Urban Health Insurance Reform in China", a presentation by Ling Xu of China's MOH in December 2004 at an MOH seminar on the 2003 NHS. The rural figures for 2003 are from "Main Findings from the 3rd NHS Survey", available online at http://www.moh.gov.cn, accessed on 21 April 2005.

the new BMI program, covering over 109 million urban employees. China's smaller cities are still in the process of implementing these reforms.

3. Methods, data and descriptive statistics

We model the effect of insurance on the probability of an individual incurring large annual out-of-pocket payments, with large being defined relative either to local average income or the household's own per capita income. We also estimate the impact of insurance on actual out-of-pocket expenses, and in an effort to shed light on the reasons for our results also report estimates of the effect of insurance on use of services in general, the public and private sectors, and the level of facility used (village, township or higher-level).

As indicated in Section 1, a fundamental estimation issue – whatever the outcome variable – is the likely endogeneity of insurance. Selection into the scheme is likely to be nonrandom, and insurance and the error term in our estimating equation are likely to be correlated with one another. This will bias the coefficient on insurance in the out-of-pocket spending equation upwards if selection into the scheme is adverse to the insurer: people with unobservables that predispose them to higher-than-expected out-of-pocket payments are more likely to enroll than others, and this may explain (at least in part) any observed positive association between insurance and out-of-pocket payments. Favorable selection, by contrast, will bias the coefficient on insurance in the out-of-pocket spending equation downwards, and a negative association between insurance and out-of-pocket payments may reflect at least in part the fact that those insured have a predisposition to lower out-of-pocket payments that cannot be captured entirely by observable variables.

With panel data (two of our three surveys are panels, one of which has four waves), we could try to take into account endogeneity through a fixed effects (FE) model, the assumption being that any correlation between the error and insurance status is due entirely to a correlation between time-invariant unobservables and insurance. This assumption, which is the identification strategy employed in the aforementioned study of Vietnam by Sepehri et al. (2006), may be unwarranted, since there may be time-varying unobservables that are also correlated with insurance. The method of instrumental variables (IV) is an obvious alternative strategy for both panel and cross-section datasets (the *only* alternative in the latter case) and is a possible *supplemental* strategy in the case of the panel datasets. If we can find instruments that are uncorrelated with the (composite) error in our estimating equation, we can simply use IV on the pooled panel. IV is available for the probit model, which is the obvious model for our high and catastrophic spending variables, as well as for whether people used services and used the private sector. For actual out-of-pocket expenses, the favored model until recently has been a two-part model, due to the large fraction of a typical sample that records zero values and the heavy right-hand tail. Recently, however, increasing use has been made of the generalized linear model (GLM) for modeling out-of-pocket expenses. Buntin and Zaslavsky (2004) found in their application that a GLM with a log link function and a variance proportional to the mean did particularly well in terms of predictive performance. This special case of the GLM is, in fact, identical to the Poisson model commonly used in the health literature to analyze count data (cf. Mullahy, 1997).[4] IV estimators are available for the Poisson model (Mullahy, 1997; Hardin et al., 2003), and indeed other GLM models too.[5] Mullahy (1998) notes, in fact, that the scope for using IV is an additional attraction of the GLM approach over the traditional two-part model. IV is not, however, feasible for the ordered probit model, which is the obvious choice for modeling the level of facility used; in this case, we have to resort to a standard IV model estimated on the pooled panel.

In all cases, we need to estimate the standard errors in a way that takes into account the likely autocorrelation in the error structure due to the panel nature of the pooled data, and which would otherwise result in unjustifiably small standard errors and hence unjustifiably large t-statistics (cf. Bertrand et al., 2004; Cameron and Trivedi, 2005, p. 705 ff.). This can be avoided by adjusting the standard errors in the IV probit and GLM models for clustering at the level of the individual observation. If it turns out there is no evidence of endogeneity, we can reasonably re-estimate the

[4] It can be verified that identical estimates are obtained using Stata's Poisson routine and its GLM routine with a log link function and a Poisson variance.

[5] We use Hardin et al.'s (2003) qvf Stata routine for estimation, which provides for quasi maximum likelihood IV estimation of GLMs, including the Poisson model, but also other GLMs considered by Buntin and Zaslavsky (2004), such as the case where the variance is proportional to the mean squared (the gamma model).

relevant models as a standard probit on the pooled sample, again ensuring that the standard errors are estimated taking into account the likely autocorrelation in the data, using the cluster option in the probit and GLM models, specifying the individual as the unit at which clustering occurs.

If we cannot find instruments that are uncorrelated with the (composite) error in our estimating equation, IV on the pooled panel is not an option. A standard FE model is one obvious fallback option in this case. This is an option in the case of the Poisson/GLM model, but not for the probit model; what we can do in this case is to estimate a FE version of the logit model. For all FE models, we need to take care to allow for possible autocorrelation in the errors, which we do via bootstrapping, with the individual as the stratum variable to ensure that samples are drawn independently from each panel member (cf. Cameron and Trivedi, 2005, p. 708). Of course, there is, as mentioned above, the possibility that the error in the FE estimating equation is correlated with insurance; i.e. there are time-varying observables correlated with insurance coverage. In a standard regression setup, one option would be to try to find instruments that are uncorrelated with the error after the estimating equation has been transformed via the within transformation, and then use IV in an FE model (cf. Cameron and Trivedi, 2005, p. 757). This cannot be done for the probit, logit or GLM models, however, and the question would be whether it is worse in this scenario to stick with models that capture the special nature of the dependent variable but only partially capture the endogeneity of insurance or to opt for a standard regression model that does not capture the special nature of the dependent variable but fully captures the endogeneity of insurance.

Our surveys are listed in Table 1. They vary in several important dimensions. One is geographic coverage: the Gansu Survey of Children and Families (GSCF) covers just Gansu province, China's second poorest province; the China Health and Nutrition Survey (CHNS) covers mostly central and eastern provinces but is a reasonably representative sample of this part of China; the World Bank Health VIII project baseline survey (H8BS), by contrast, covers mostly China's poorer central and western provinces. Another difference is the rural–urban focus within provinces: the GSCF and the H8BS are rural-only surveys, while the CHNS covers urban and rural areas. A third difference is the panel dimension: the CHNS and GSCF are both panels (we use 4 waves of the CHNS and both currently available waves of the GSCF), while the H8BS is a cross-section. A final obvious difference concerns the timing of the survey: the CHNS data were collected in 1991, 1993, 1997 and 2000; the GSCF data were collected in 2000 and 2004; and the H8BS data were collected in 1998. There are other less obvious differences, as will become apparent below.

Variable definitions and notes on the variables are provided in Table 2. Out-of-pocket health spending is for a 4-week window in the case of the CHNS (even for inpatient care), and for 1-year windows in the two other surveys. In the CHNS, out-of-pocket expenses are explicitly net of any reimbursement that patients have received or expect to receive; in the other surveys, respondents are not explicitly asked about to report expenses net of reimbursement. A further difference is that the CHNS out-of-pocket spending data refer to the individual, while in the other two surveys the data are collected at household level and then converted to a per capita basis. Finally, it is worth noting that health expenses in the CHNS are constructed from a series of follow-up questions about different types of utilization after a question asking whether people had been sick or injured during the previous 4 weeks, whereas in the other two surveys expenses were obtained from a single question on (household) health expenses.

We define expenses as 'high' if they exceed a given percentage threshold of mean per capita household income in the household's community or primary sampling unit (a village or urban district), and as 'catastrophic' if they exceed a given percentage of the household's own per capita income. In each case, we present results for five thresholds: 5%, 10%, 15%, 20% and 25%.[6] As is apparent from Table 3, the rates reporting 'high' and 'catastrophic' out-of-pocket payments vary considerably across the surveys (the CHNS having far lower rates), and to a lesser extent over time within each of the two longitudinal surveys. The far lower rates in the CHNS reflect in part the relatively low out-of-pocket payment amounts in our data (they refer to a 1-month window, but even so they are lower when annualized

[6] Wagstaff and van Doorslaer (2003), in their study of catastrophic spending in Vietnam, use thresholds of 2.5%, 5%, 10% and 15%, when income is the denominator as in the present study, but also present results for catastrophic spending defined relative to discretionary income (income less food spending), where 10%, 20%, 30% and 40% as thresholds. Xu et al. (2003), in their international comparative study, use a threshold of 40% with a denominator of income less an allowance for food spending. In their study of the US, Shen and McFeeters (2006) define out-of-pocket payments as being a high burden if they exceed 5% of family income, the threshold used in the US government's State Children's Health Insurance Program. Waters et al. (2004) use a figure of 10% in their US study. Van Doorslaer et al. (2007) in their pan-Asia comparisons use the same thresholds for catastrophic spending as we use here.

Table 1
Surveys used

	China Health and Nutrition Survey	Gansu Survey of Children and Families	World Bank China Health VIII Project Baseline Survey
Provinces sampled	Guangxi, Guizhou, Heilongjiang (1997 and 2000 only), Henan, Hubei, Hunan, Jiangsu, Liaoning (not 1997), Shandong	Gansu (1)	Anhui, Chongqing, Gansu, Guizhou, Henan, Qinghai, Shanxi
No. of counties sampled	36	20	28
Rural or urban or both	Both	Rural	Rural
Areas covered	Where possible, the provincial capital and a lower-income city were selected, though in two provinces, other large cities had to be selected. Villages and townships within the counties, and urban and suburban neighborhoods within the cities were sampled.	Villages only	Villages and townships
Sampling strategy	A multistage, random cluster process was used to draw the sample surveyed in each of the provinces, which vary substantially in geography, economic development, public resources, and health indicators. Counties in the nine provinces were stratified by income (low, middle, and high) and a weighted sampling scheme was used to randomly select four counties in each province. Villages and townships within the counties and urban and suburban neighborhoods within the cities were selected randomly[a]	A four-stage stratified random sample was used. First counties were selected (20), then townships (42), then villages (100) and finally children. At each stage a sampling procedure was used to ensure that sampling was done evenly across the income distribution[b]	Survey was administered only in project counties[c] which were deliberately selected on the basis of their high rates of poverty, their capacity to implement the project, and their financial resources (counties were responsible for repaying the loan to the World Bank). Within counties, a random cluster process was used to draw households
Date(s) conducted	1991, 1993, 1997, 2000[d]	2000, 2003	1998
Sample size (no. individuals) per wave	14,578; 13,687; 14,181; 15,334	7581, 7581	18,200. Individuals aged less than 20 excluded from estimation sample

[a] Details taken from CHNS website http://www.cpc.unc.edu/projects/china.
[b] Further details are to be found at the GSFC website http://www.ssc.upenn.edu/china/gscf/mainGscf.htm.
[c] The survey covers 22 of the 96 counties eventually included in the Health VIII project. The 22 counties are located in the following provinces in central and western China: Anhui, Chongqing, Gansu, Guizhou, Henan, Qinghai and Shanxi.
[d] The CHNS also collected data in 1989. This wave was excluded from the present analysis because some of the variables were not collected in the 1989 wave.

than those in the other surveys).[7] In part, this likely reflects the 4-week reporting period in the CHNS and the fact that the spending data were collected through a series of questions put to household members who had been ill or injured during the preceding 4 weeks. Those in hospital at the time of the interview were presumably not interviewed at all. With an average length of stay in China of 12 days in 2000 and 15 days in 1990, the CHNS likely captures only a part of household inpatient-related expenses in the 4 weeks prior to the survey. We have trimmed outliers on the out-of-pocket

[7] It seems likely that the short (4-week) recall period used in the CHNS would result in a larger annualized out-of-pocket spending figure than the long (12-month) recall period used in the other two surveys. This is borne out by analysis of data from the 1998 Vietnam Living Standards Survey, which inquires about health spending over a 12-month period, and about use of services and medical expenses during the last 4 weeks. Responses to the spending questions covering the last 12 months produce a budget share (share of per capita household consumption) of 5.2%, while the questions covering the last 4 weeks produce an annualized budget share of 9.1%. The fractions of the sample with 'high' and 'catastrophic' expenses, defined along the lines of the present paper, are also higher when the 4-week questions are used.

Table 2
Variable definitions

	China Health and Nutrition Survey	Gansu Survey of Families and Children	World Bank China Health VIII Project Baseline Survey
Out-of-pocket health spending	Total health care expenditures during last month, including expenditures not associated with a provider visit, expenditures associated with first visit to providers (if applicable), expenditures associated with visit to second provider, additional health care expenditures, minus expenditure reimbursed by health insurance (1989 prices). The top half of 1% of observations in each wave has been trimmed, and preventive expenditures have been excluded	Sum of expenditure associated with doctor visits and drug purchases during last 12 months	Annual expenses on drugs, prevention and health care of the family. The top half of 1% of observations in each wave has been trimmed
High out-of-pocket spending	Last month's out-of-pocket spending in excess of *x*% of mean monthly per capita income in the household's primary sampling unit for the wave in question. *x* defined as 5, 10, 15, 20 and 25	Annual out-of-pocket spending in excess of *x*% of mean per capita income in the household's primary sampling unit for the wave in question. *x* defined as 5, 10, 15, 20 and 25	Annual out-of-pocket spending in excess of *x*% of mean per capita income in the household's primary sampling unit for the wave in question. *x* defined as 5, 10, 15, 20 and 25
Catastrophic out-of-pocket spending	Last month's out-of-pocket spending in excess of *x*% of household's own monthly per capita income for the wave in question. *x* defined as 5, 10, 15, 20 and 25	Annual out-of-pocket spending in excess of 10% of household's own per capita income for the wave in question. *x* defined as 5, 10, 15, 20 and 25	Annual out-of-pocket spending in excess of 10% of household's own per capita income. *x* defined as 5, 10, 15, 20 and 25
User	Binary variable equaling one if person used any health facility in last 4 weeks		Binary variable equaling one if person used any health facility in last 4 weeks
Private	Binary variable equaling one if person used private clinic in last 4 weeks		Binary variable equaling one if person used private clinic in last 4 weeks
Provider level	0 = none, 2 = village clinic, 3 = township health center, 4 = hospital		0 = none, 2 = village clinic, 3 = township health center, 4 = hospital
Health insurance	A dummy indicating whether the person is covered by any health insurance scheme	A dummy indicating whether the person has any health insurance cover	A dummy indicating whether the person was covered by GIS/LIS schemes (people saying they had one of the following: government employee insurance; labor insurance; half labor insurance; medical insurance; coordinated arrangement; others). CMS membership was included separately, and defined at the village level (only a few villages had a CMS, and where there was any membership, it was over 90%)

Per capita income	Total household income from all sources divided by number of household members (1989 prices). The bottom half of 1% of observations in each wave have been trimmed. Total household income has been constructed by a colleague of the authors from the original data on various income components in favor of the CHNS 'official' income aggregate which has been questioned by researchers and which fails to capture the large rise in per capita incomes over the period covered by the CHNS. Income has been converted to 1989 prices using the CHNS price indices	Income from agriculture, livestock, wages and self-employment. Wage income includes bonuses, subsidies, and the value of in-kind payments. The bottom half of 1% and top 1% of observations in each wave have been trimmed, the latter because the unadjusted data showed an implausibly high rise in income between 2000 and 2004 compared to data for Gansu from the government's Rural Household Survey. Income has been converted to real terms using the China rural CPI	Household was asked its total household income in the previous year. This was divided by the number of household members to get per capita income. The bottom half of 1% of observations in each wave have been trimmed
Health	4-Point self-assessed health scale	5-Point self-assessed health scale	A dummy indicating whether the individual has a diagnosed chronic disease
Education	Years of education	Years of education	Education in categories: no education (omitted category), elementary school, primary middle school, senior middle school, vocational school, junior college
Age	Age in brackets: 0/45, 46/65, 66+	Age in years	Age in years and age squared
Other covariates	Urban/rural. Wave of panel	Child. Pensioner. Relation to focus child of the GSCF (parent, sibling, etc.). CMS in village. Employed. Perceived quality of care in village clinic. No. doctors in village. No. nurses in village clinic. No. midwives in village. Villager's preferred provider	Male. No. children in household. Age of oldest household member. Education of household head. Employment of household head. Village income per capita. Electricity in village. Road in village. Telephone in village. Percentage of households in village covered by 'five-guarantees' anti-poverty program. Percentage of households in village poor as defined by 640-RMB poverty line. Distance from village to county hospital. Whether village clinic operates essential drugs list
Instruments for insurance	Government official. Head of household. Working member of household	Perceived quality of care in local township health center	Employment status, categories being: worker, Peasant, skilled worker, student, child, housework, unemployed, retired, and other

Table 3
Descriptive statistics for outcomes and insurance, by survey and wave

	CHNS				GSCF		H8BS
	1991	1993	1997	2000	2000	2003	1998
Actual expenses	1.678	0.916	1.739	2.513	108.223	140.033	78.062
Catastrophic (5%)	0.042	0.020	0.034	0.034	0.751	0.654	0.633
Catastrophic (10%)	0.034	0.017	0.028	0.027	0.659	0.522	0.352
Catastrophic (15%)	0.029	0.014	0.023	0.024	0.585	0.448	0.224
Catastrophic (20%)	0.025	0.013	0.020	0.021	0.525	0.396	0.143
Catastrophic (25%)	0.023	0.012	0.019	0.019	0.479	0.355	0.102
High (5%)	0.041	0.020	0.031	0.032	0.754	0.649	0.624
High (10%)	0.032	0.015	0.024	0.024	0.659	0.500	0.350
High (15%)	0.025	0.013	0.020	0.020	0.572	0.418	0.199
High (20%)	0.022	0.011	0.017	0.017	0.498	0.360	0.124
High (25%)	0.019	0.010	0.014	0.016	0.444	0.318	0.084
User	0.051		0.036	0.037			0.048
Private clinic use	0.003		0.005	0.007			0.006
Facility type	0.104		0.065	0.068			0.070
Insurance	0.270	0.231	0.238	0.200	0.003	0.180	0.037

Note: Out-of-pocket spending figure for CHNS is for last month; for other surveys refers to last 12 months.

payments distributions: the top half of one percent of cases on the out-of-pocket payments distribution was dropped. Despite the differences in rates, one common feature of the three surveys is the positive correlation among the various high and catastrophic spending variables, and their negative correlation with income—in China, it is the worse off who are more likely to experience catastrophic health expenses.[8]

We define health insurance in the CHNS and GSCF as any coverage, while in the H8BS we have focused on GIS/LIS membership.[9] Insurance coverage varies across the surveys and across waves in the case of the longitudinal surveys. The higher coverage in the CHNS reflects the fact the survey – unlike the other two surveys – covers urban as well as rural households. The trend in the CHNS is downwards, which is broadly in line with other survey data (the government's 5-yearly National Health Survey showed a decline between 1993 and 1998, and a slight upturn between 1998 and 2003). In the case of the GSCF, the rise in coverage between 2000 and 2004 seems to be due to an expansion of insurance coverage among children for which the contributions (and presumably the benefits) are fairly low. In the H8BS, there are a couple of villages that still had a CMS and coverage in these villages was universal. We have included the presence of a CMS as a covariate in the H8BS analysis, and have treated it as exogenous. It is worth emphasizing that insurance coverage refers to the individual while out-of-pocket payments in the GSCF and H8BS refer to the entire household. We would expect to see a closer link between insurance and out-of-pocket payments in the CHNS than in the GSCF and H8BS.

We include several common covariates in the estimating equations for all surveys, namely per capita income, health, education and age. Information on income is most complete in the CHNS, the GSCF has the second most detailed income information, while income in the H8BS is based on just one question. Further details of the income variables are contained in Table 2.[10] Two of the surveys contain self-assessed health (in the GSCF, the assessment was actually done by the woman of the household). In the case of the H8BS, no self-assessment of health is available, and we used instead a chronic illness dummy. We have measured schooling by years of schooling in the CHNS and the GSCF, but in the H8BS have left education coded as levels of attainment.

[8] This emerges also in the analysis of Van Doorslaer et al. (2007). In most other Asian countries, this is not the case.

[9] In the CHNS, 85% of those who have health insurance are in the GIS or LIS scheme. Even in rural areas, these are the commonest schemes, the other schemes being very uncommon. We explored the implications of defining health insurance as GIS/LIS and including other insurance as an exogenous covariate. The results, which are among the supplementary results available on request from the authors, were very similar.

[10] Income is notoriously hard to measure accurately in household surveys in developing countries. To get a sense of how sensitive our results might be to measurement error, we re-estimated our CHNS models using the 'official' CHNS aggregate income variable, which has been questioned by researchers, not least on the grounds that it fails to reflect the rapid growth of income in China over the period covered by the panel. The findings were surprisingly robust, perhaps due in part to our use of time dummies which may be picking up differences across the waves in the way the income data were collected. Results are available from the authors on request.

In addition to the common covariates, we have included a number of survey-specific covariates, listed in Table 2. In the two panels, we included panel-specific fixed effects. In part, the choice of other covariates to include was decided in part according to the results of preliminary over-identification tests, since inevitably whether or not a variable could be legitimately excluded from the outcome equation depended in part on what variables were included in the outcome equation as well as insurance status. One additional factor was that out-of-pocket spending is measured at the household level in the GSCF and H8BS, which meant it was important we try to capture relevant data on other household members as well as the individual in question.

Our instruments for insurance vary across the three surveys, and our final choice was the result of an informed process of trial and error. We use formal tests to guide us, but also build on what we know about the schemes, namely that in the CHNS and H8BS the schemes in question are largely formal-sector insurance schemes, while in the GSCF the scheme seems to be geared largely to school children. In the CHNS we use three variables as instruments: whether the individual is a government official; whether he or she is head of the household; and whether the individual is a working member of the household. In the GSCF, perceived quality of the local township health center, measured on a 3-point scale and in the eyes of a key informant. In the H8BS, we use a vector of variables capturing employment status. The assumption is that our instruments are correlated with the probability of the individual being insured but are not correlated with out-of-pocket spending and the probability of high or catastrophic spending only through insurance once we have controlled for health insurance and our other covariates, such as household income, education, and so on. Our instruments all pass the "relevance" test (they are correlated with insurance) and typically also pass the joint test of validity (they are uncorrelated with the error term) and appropriate exclusion (they can justifiably be excluded from the outcome equation).

4. Results

Our main results are reported in Tables 4–6. We report coefficient estimates only for the insurance variable (results for other variables are available on request from the authors), but in all cases the full set of covariates listed in Table 2 are included in the model. The first results are for the regular GLM/Poisson and probit models, estimated in the case of the two panels on the pooled data with standard errors adjusted for autocorrelation. Each table then reports the

Table 4
CHNS main results

	Poisson/probit		ALN over-identification test (prob)	Wald test of endogeneity (prob)	IV Poisson/probit		FE Poisson/logit	
	Coefficient	*t*-Statistics			Coefficient	*t*-Statistics	Coefficient	*t*-Statistics
Actual expenses	0.104	0.92			0.571	1.03	0.194	0.92
Catastrophic (5%)	0.050	1.42	0.723	0.070	0.380	2.06	0.248	1.31
Catastrophic (10%)	0.059	1.52	0.397	0.124	0.368	1.79	0.304	1.71
Catastrophic (15%)	0.053	1.27	0.569	0.174	0.342	1.56	0.337	1.74
Catastrophic (20%)	0.038	0.84	0.562	0.451	0.211	0.87	0.273	1.33
Catastrophic (25%)	0.042	0.89	0.634	0.469	0.213	0.85	0.295	1.43
High (5%)	0.038	1.11	0.624	0.114	0.331	1.76	0.315	2.14
High (10%)	0.009	0.22	0.471	0.084	0.355	1.74	0.221	1.28
High (15%)	0.000	−0.01	0.639	0.274	0.244	1.08	0.215	1.03
High (20%)	−0.016	−0.36	0.687	0.502	0.148	0.60	0.226	0.98
High (25%)	−0.004	−0.08	0.986	0.413	0.205	0.79	0.287	1.31
User	0.166	4.90	0.295	0.392	0.327	1.71	0.522	2.58
Private clinic use	−0.460	−4.48	0.182	0.938	−0.527	−0.86	−1.092	−0.15

Notes: Table shows coefficient estimates and *t*-statistics for insurance variable in model that includes all other covariates listed in Table 2. GLM/Poisson is used to model actual expenses, probit for all other outcomes in table. Estimates for Poisson/probit and IV Poisson/probit are obtained from pooled panels for 1991, 1993, 1997 and 2000. All *t*-statistics including those from the FE Poisson and logit models are based on standard errors that are robust to autocorrelation and heteroskedasticity, obtained via bootstrapping with individual ID as the strata variable. ANL test is Amemiya–Lee–Newey over-identification test, the null of which is that the instruments are valid (i.e. uncorrelated with the error term) and correctly excluded from the outcome equation; a rejection casts doubt on the validity of the instruments. The null hypothesis of the Wald test of endogeneity is that insurance can be treated as exogenous; a rejection points to insurance being endogenous.

Table 5
GSCF main results

	Poisson/probit		ALN over-identification test (prob)	Wald test of endogeneity (prob)	IV Poisson/probit		FE Poisson/Logit	
	Coefficient	*t*-Statistics			Coefficient	*t*-Statistics	Coefficient	*t*-Statistics
Actual expenses	0.335	3.06			0.925	0.78	0.447	3.35
Catastrophic (5%)	0.158	3.75	0.010	0.000	2.781	9.88	0.162	1.34
Catastrophic (10%)	0.199	4.74	0.061	0.000	2.616	8.58	0.279	2.49
Catastrophic (15%)	0.187	4.37	0.054	0.000	2.406	6.86	0.237	2.13
Catastrophic (20%)	0.172	3.94	0.144	0.000	1.985	4.72	0.198	1.71
Catastrophic (25%)	0.202	4.52	0.736	0.001	1.891	4.42	0.217	1.87
High (5%)	0.145	3.45	0.016	0.000	2.429	6.76	0.164	1.49
High (10%)	0.132	3.19	0.021	0.000	2.487	7.51	0.178	1.52
High (15%)	0.182	4.33	0.339	0.006	1.622	3.45	0.248	2.36
High (20%)	0.206	4.76	0.257	0.016	1.521	3.05	0.305	3.04
High (25%)	0.195	4.40	0.064	0.146	1.066	1.85	0.320	2.87

Notes: Table shows coefficient estimates and *t*-statistics for insurance variable in model that includes all other covariates listed in Table 2. Poisson is used to model actual expenses, probit for all other outcomes in table. Estimates for Poisson/probit and IV Poisson/probit are obtained from pooled panels for 2000 and 2004. All *t*-statistics including those from the FE Poisson and logit models are based on standard errors that are robust to autocorrelation and heteroskedasticity, obtained via bootstrapping with individual ID as the strata variable. For notes on ANL and Wald tests, see Table 4. In IV Poisson model, variable indicating whether village had CMS had to be dropped from regressor set to achieve convergence.

Table 6
H8BS main results

	Poisson/probit		ALN over-identification test (prob)	Wald test of endogeneity (prob)	IV Poisson/probit	
	Coefficient	*t*-Statistics			Coefficient	*t*-Statistics
Actual expenses	0.086	1.22			0.446	2.72
Catastrophic (5%)	−0.004	−0.04	0.154	0.054	0.444	1.97
Catastrophic (10%)	0.036	0.32	0.214	0.106	0.420	1.72
Catastrophic (15%)	−0.117	−0.92	0.483	0.002	0.598	2.46
Catastrophic (20%)	−0.136	−1.00	0.106	0.011	0.506	1.74
Catastrophic (25%)	−0.036	−0.25	0.021	0.019	0.681	1.98
High (5%)	−0.031	−0.31	0.072	0.120	0.342	1.47
High (10%)	−0.033	−0.34	0.674	0.030	0.399	2.00
High (15%)	0.045	0.47	0.048	0.036	0.470	2.39
High (20%)	0.065	0.58	0.042	0.037	0.495	2.18
High (25%)	0.181	1.59	0.016	0.022	0.717	2.76
User	0.149	1.04	0.491	0.012	0.743	2.73
Private clinic use	−0.070	−0.32	0.729	0.232	0.968	1.23

Notes: Table shows coefficient estimates and *t*-statistics for insurance variable in model that includes all other covariates listed in Table 2. Poisson is used to model actual expenses, probit for all other outcomes in table. For notes on ANL and Wald tests, see Table 4. In IV Poisson model, variable indicating number of children had to be specified as endogenous/subject to measurement error to achieve convergence.

Amemiya–Lee–Newey (ALN) over-identification test, a rejection of which would cast doubt on the validity of our instruments. (In all cases we were able to reject the hypothesis that the instruments are irrelevant: for all datasets, we obtained a zero probability value of the Wald statistic testing the hypothesis that the coefficients on the instruments in the first-stage.) The tables then report the probability value of a Wald test of the hypothesis that insurance is exogenous, a rejection of which points to insurance being endogenous, and IV estimates for the Poisson and probit models which are the relevant estimates in cases where the hypothesis of instrument validity is not rejected and the hypothesis of exogeneity is rejected. Also reported are FE Poisson and logit estimates, which are especially relevant in the case where our instruments are not valid.[11]

[11] Over-identification and endogeneity tests are not reported for out-of-pocket payments. Mullahy (1997) uses a Hansen-type J-statistic for his GMM IV routine, but our results are obtained via quasi maximum likelihood.

We discuss first the impacts of insurance on out-of-pocket spending and the incidence of high and catastrophic spending, coming back to the effects of insurance on the use of private providers later. In the CHNS results in Table 4 the instruments emerge as valid according to the ANL test for all spending variables. At the 1% significance level, insurance emerges as exogenous for all outcomes. However, at the 5% level insurance is endogenous for catastrophic expenses at the 5% threshold. Insurance is borderline endogenous at the 10% significance level for catastrophic expenses at the 10% threshold and for high expenses at the 5% threshold. While none of the estimates suggests that insurance significantly increases out-of-pocket expenditures, several of the results for the risk of catastrophic and high out-of-pocket spending point to health insurance increasing financial risk. According to the IV probit estimates, insurance significantly increases (at the 10% significance level or better) the risks of catastrophic expenditures at the 5% and 10% thresholds and the risk of high expenditures at the 5% threshold (the thresholds where insurance is endogenous or borderline endogenous). Broadly similar conclusions emerge from the FE results as from the IV probit results. At other thresholds, insurance has no effect on the risk of catastrophic or high expenses for any of the estimation methods.

The GSCF results in Table 5 present a more clear-cut picture: irrespective of the estimation method, insurance significantly increases actual out-of-pocket payments (except in the case of the IV estimates) and typically the risk of catastrophic and large expenses as well. The instruments work somewhat less well in this dataset, but pass the over-identification test in all but three outcomes. Insurance emerges as endogenous except in the case of high expenses at the 25% threshold, and the relevant estimates (typically IV) point to a statistically significant effect of insurance on actual expenses and the risk of high and catastrophic spending. The FE results mostly reinforce the IV results, and point to insurance increasing actual expenditures.

The H8BS results in Table 6 suggest we can have a good deal of confidence in our instruments for the majority of outcomes, though for four outcomes the null hypothesis of over-identification is rejected at the 5% level (although in one case only marginally). Where the instruments are valid, the hypothesis of insurance being exogenous is usually rejected, and the (IV) results imply that insurance significantly increases the outcome in question, namely the risk of catastrophic spending at all but the 25% threshold. For this latter outcome and for the high expense variables, it is less clear whether our instruments are valid, but the exogeneity tests suggest insurance is endogenous and the IV results point to significant impacts of insurance. The IV results also point to insurance raising actual out-of-pocket payments.

Taken together, the results from the three surveys suggest that health insurance in China *raises* rather than reduces out-of-pocket spending and the risk of catastrophic and large expenses. The supplemental results in Tables 4 and 6 showing the effect of insurance on the probability of using services and of using a private clinic, as well as those in Tables 7 and 8 showing the effect on the level of provider chosen, give some clues as to why this occurs. Our instruments in the any-use equation are valid in Tables 4 and 6, with insurance apparently being exogenous in the CHNS but endogenous in the H8BS. The relevant estimates suggest that insurance increases the probability of people using health services. This could be due to insurance increasing rates of illness, but is more likely to reflect people

Table 7
Basic results for level of provider

	Coefficient	*t*-Statistics	Weak identification test (*F*)	Under-identification test statistic (prob)	ALN over-identification test (prob)	Wald test of endogeneity (prob)
CHNS						
OLS	0.044	7.28				
IV	0.114	2.75	293.74	0.000	0.042	0.075
FE	0.031	3.09				
IV/FE	0.325	1.59	15.33	0.000	0.182	0.133
H8BS						
OLS	0.023	0.72				
IV	0.117	2.49	21.45	0.000	0.242	0.137

Notes: See Table 2 for definition of dependent variable. Table shows coefficient estimates and *t*-statistics for insurance variable in model that includes all other covariates listed in Table 2. *t*-Statistics for CHNS sample are based on standard errors that are robust to autocorrelation and heteroskedasticity, while *t*-statistics for H8BS are based on standard errors that are robust to clustering at village level and heteroskedasticity. Weak identification test statistic is Kleibergen-Paap rk Wald *F* statistic. The null hypothesis is that excluded instruments are correlated with the endogeous regressors, but only weakly. Underidentification test statistic is Kleibergen-Paap rk LM statistic. The null hypothesis is that the equation is under-identified, i.e., that the excluded instruments are not "relevant", meaning that they are uncorrelated with the endogenous regressors. For further details on weak and under-identification tests, see Baum et al. (2007). For notes on ANL and Wald tests, see Table 4.

Table 8
Ordered probit results for level of provider

	Incl. non-users		Excl. non-users	
	Coefficient	z-Statistics	Coefficient	z-Statistics
CHNS				
Coefficient	0.206	6.16		
Marginal effects				
No facility	−1.09E−04	−5.10		
Village clinic	7.00E−05	5.18	−0.252	−8.96
THC	2.04E−05	4.81	−0.013	−2.14
Hospital	1.83E−05	4.51	0.266	8.28
H8BS				
OLS	0.023	0.72		
IV	0.117	2.49		
Coefficient	0.146	1.00		
Marginal effects				
No facility	−1.34E−02	−0.90		
Village clinic	7.93E−03	0.90	0.025	0.21
THC	4.17E−03	0.91	−0.016	−0.20
Hospital	1.30E−03	0.84	−0.009	−0.21

See Table 2 for definition of dependent variable. In first pair of columns, non-users (coded zero on provider type variable) are included. They are excluded in second pair of columns, where the focus is on provider level among users of some facility. Coefficient shows effect of insurance on the underlying latent variable. Small marginal effects in the first model reflect the preponderance of non-users in the sample.

being more inclined to seek care when they fall ill. In both the CHNS and the H8BS our instruments in the case of use of a private clinic are valid, and in both cases insurance emerges as exogenous. In both surveys insurance has a negative effect on the likelihood of people using a private clinic. This likely reflects two factors: the fact that care received at a private clinic is likely not to be covered, and the fact that once they are insured people use higher-level providers (private clinics are village clinics, and the lowest level of provider).

The effect of insurance on level of provider is examined in Tables 7 and 8 (see Table 2 for definition of this variable). In this case, the favored model would be the ordered probit, were it not for our concern over endogeneity, there being no IV or FE versions of the ordered probit model. Table 7 ignores the ordinal nature of the provider level variable, and runs the standard over-identification and endogeneity tests. The instruments for the CHNS data pass the over-identification test more convincingly in the FE specification than in the levels specification, where insurance emerges as exogenous. The standard FE results would seem therefore the most reliable, though they are actually no different qualitatively from the OLS and IV results on the pooled data: all have a positive and significant coefficient on insurance, which would seem to point to insurance pushing people toward higher-level providers. This is borne out by the ordered probit results for the CHNS subsample of users: the marginal effects in Table 8 point to insurance reducing the likelihood of users receiving care from village clinics and THCs, and increasing the likelihood of them receiving care from hospitals. When the model is estimated on the CHNS sample of users and non-users (with 'no provider' being the base category), the marginal effects are non-monotonic, but this model confounds the effect of insurance on the decision to seek care and its impact on the level of provider chosen conditional on care being sought. The H8BS results are less clear-cut than the CHNS results. Insurance is exogenous at the 10% level (but not at the 15% level) in Table 7, but only in the IV results is the insurance effect statistically significant. None of the marginal effects in the ordered probit are significant. The difference between the CHNS and H8BS results may reflect the rural nature of the H8BS sample and the mixed rural–urban nature of the CHNS sample, as well as the fact that the H8BS sample is poorer.

5. Summary and conclusions

The results in this paper are consistent with the hypothesis that health insurance need not always reduce financial risk. Indeed, our results from three separate household surveys suggest that in China health insurance is more likely than not to *increase* out-of-pocket spending and to *increase* the risk of catastrophic and large expenses. Our results

do not, we believe, reflect the endogeneity of insurance. They do not reflect, in other words, people with unobserved characteristics that predispose them to high spending being more likely to opt for health insurance. We have assembled instruments for health insurance that are relevant and (mostly) valid according to standard over-identification tests for limited dependent variable models. Where we have discovered evidence pointing to insurance being endogenous, we have used instrumental variable methods, but have also reported fixed effects model estimates; in one case we have combined the two. Interestingly, when we take into account the endogeneity of insurance, our estimates typically increase rather than fall, suggesting that there is any selection on unobservables, it is *favorable* to the scheme rather than adverse to it.

Our supplemental results give some clues as to why health insurance in China does not reduce financial risk. They suggest that insurance makes people more likely to use health care providers, we hypothesize by making them more inclined to seek care when they fall sick rather than making them more likely to fall sick. The supplemental results also suggest that insurance makes people more likely to move up the provider 'ladder': preferring township health centers (THCs) to village clinics, and hospitals to THCs. These results suggest that at least in China health insurance produces discrete shifts in utilization patterns, and that as a result out-of-pocket payments may exhibit lumpiness, with use versus non-use producing one discrete jump in spending, and shifts to successively higher-level providers producing further discrete jumps. This is not to say that insurance in China may not also be associated with increases in out-of-pocket payments *at a given level of provider*, with hospitals, say, delivering more costly tests, drugs and medical interventions to people who have insurance coverage.[12] There may be lumpiness here too if care is subject to indivisibilities; this seems especially likely to be the case at higher-level facilities where high-tech care may become an option once people have insurance, and less likely at lower levels of care where insurance coverage may raise out-of-pocket spending by people being prescribed more drugs and more expensive drugs. We leave to future studies the task of ascertaining how far the additional spending associated with insurance reflects (a) changes in spending as insured patients shift up the provider ladder and (b) increases in out-of-pocket payments at a given rung in the provider ladder.

While our results make clear that expanding insurance coverage does *not* (contrary to what policymakers and observers often think) automatically improve financial protection in health, their welfare implications are not clear-cut. People would like to weigh any extra risk of large out-of-pocket payments against the additional health gains from being able to receive more extensive and more sophisticated medical care once insured. On balance, they may be better off with insurance despite facing a higher financial risk. If, however, providers exploit their informational advantage and take the opportunity of insurance coverage to deliver expensive medical care that the individual would not have chosen had he been fully aware of the magnitude of the additional health benefits and additional out-of-pocket expenses, then the welfare gains associated with insurance are less clear. It is possible, in fact, that by encouraging people to seek care from higher-level providers, insurance may accentuate the informational asymmetry between provider and patient, since patients may be at a bigger informational disadvantage where high-tech care is involved than where the care involved is fairly basic. This points to the need for future research in this field to pinpoint how the care that patients receive changes as a result of their having insurance and determining the magnitude of the health gains associated with it (cf. e.g. Doyle, 2005), but also seeing whether people would, given perfect information, choose the extra care despite the extra expenses involved.

Finally, our results raise questions about the factors that influence the effect that insurance has on financial risk. The benefit package seems likely to be one factor—out-of-pocket payments and the risk of large out-of-pocket payments seem likely to depend on the size of any deductibles and ceilings, the coinsurance rate, and the range and type of services covered. But exactly how the design of the benefit package affects financial risk is unclear *a priori*: on the one hand, a more generous scheme ought, on the face of it, to reduce out-of-pocket spending and the risk of high levels of spending; but on the other hand a generous scheme may be more likely to encourage people to seek care in the first place and lure them to higher-level facilities where the information asymmetry between them and the provider may be more acute. In China, the *de jure* benefit packages differ across different schemes: the GIS, for example, has traditionally had very generous coverage, and some cities (e.g. Shanghai) are known to have more generous schemes than others. In the LIS, the *de jure* and *de facto* benefit packages started to diverge during the 1980s and 1990s, as struggling state-owned enterprises and rising health care costs coupled with small risk pools meant that the scheme was

[12] It is also possible that providers may simply charge higher prices to the insured for the same service. In China, this seems less likely given prices are regulated by the government. The adjustment seems more likely to be in terms of the type of care delivered.

unable to meet its financial obligations, one consequence being that people were dissuaded from seeking care included in the benefit package (Henderson et al., 1995). The reforms of the 1990s merged the GIS and LIS into a single scheme with a larger risk pool and a sounder financial base (which could be expected to reduce the risk of large out-of-pocket payments) but also introduced demand-side cost-sharing through the introduction of medical savings accounts (which could be expected to increase out-of-pocket spending). The net effect in Zhenjiang appears to have been to increase out-of-pocket spending among the better off but reduce it among the less well off; this seems to be due in part at least to the substitution of outpatient care for inpatient care especially among those on lower incomes (Liu and Zhao, 2006). But more research on the impacts of benefit package reform for the risk of large out-of-pocket payments is needed.

The extent of any cost-sharing on the supply side seems likely to be another factor. But here again, the likely effect on out-of-pocket payments is not clear *a priori*. In China, as in many other countries, insurers have begun to shift from FFS to some form of prospective payment. On the face of it, this ought to reduce costs overall and hence out-of-pocket payments by patients. However, providers are likely to look for ways to recoup the lost revenues they would otherwise sustain when insurers shift from FFS. Insurers themselves might be forced to pay more than warranted as providers respond to the introduction of DRGs by upcoding. If there is demand-side cost-sharing, the shift to prospective payments may in such circumstances have little impact on out-of-pocket payments. But the outcome for patients in terms of out-of-pocket payments could be worse. Faced by a tough and well-informed insurer, providers may decide that raising revenues from patients is the easier course. They might target the uninsured, raising prices or inducing further demand for their services. How far this is possible is likely to depend on whether prices are regulated, on the degree of regulation of provider activities (including self-regulation), and perhaps on market concentration (higher concentration being hypothesized to encourage greater inducement efforts). There is some evidence that hospitals in Shanghai increased revenues from uninsured patients following the introduction of DRGs by the insurer (Zhang, 2007). But providers might also try to increase out-of-pocket payments by the insured, by inducing demand for uncovered services or by engaging in extra-billing. Another challenge for future research is to examine in different settings the consequences of provider payment reform for the risk of large out-of-pocket payments among the insured (and uninsured).

Acknowledgements

Our thanks to Agbessi Amouzou, Shengchao Yu and Lei (Lydia) Liu respectively for their help preparing the China Health and Nutrition Survey data, the Gansu Survey of Children and Families (GSCF) data, and the Health VIII baseline survey data. Our thanks also to Eddy van Doorslaer, Ola Granström, Owen O'Donnell, Joe Newhouse and two anonymous referees for helpful comments on an earlier version of the paper. Wave 1 of the GSCF was funded by The Spencer Foundation Small and Major Grants Programs, while Wave 2 was funded by grants from the Fogarty International Center at the National Institutes for Health and the World Bank Research Committee. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

References

Baum, C.F., Schaffer, M.E., Stillman, S., 2007. Enhanced routines for instrumental variables/GMM estimation and testing. Boston College Working Papers in Economics 667. Boston College Department of Economics, Boston, MA.

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

Buntin, M.B., Zaslavsky, A.M., 2004. Too much ado about two-part models and transformation? Comparing methods of modeling medicare expenditures. Journal of Health Economics 23 (3), 525–542.

Cameron, A., Trivedi, R., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, Cambridge.

Culyer, A.J., 1989. The normative economics of health care finance and provision. Oxford Review of Economic Policy 5, 34–58.

Doyle Jr., J.J., 2005. Health insurance, treatment and outcomes: using auto accidents as health shocks. Review of Economics and Statistics 87 (2), 256–270.

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

Finkelstein, A., McKnight, R., 2005. What did medicare do (and was it worth it)? Working Paper 11609. National Bureau of Economic Research, New York.

Gao, J., Qian, J., Tang, S., Eriksson, B., Blas, E., 2002. Health equity in transition from planned to market economy in China. Health Policy and Planning 17 (Suppl. 1), 20–29.

Gertler, P., Solon, O., 2000. Who Benefits From Social Health Insurance in Developing Countries? University of California at Berkeley, Mimeo, Berkeley, CA.

Hardin, J., Schmiediche, H., Carroll, R., 2003. Instrumental variables, bootstrapping, and generalized linear models. Stata Journal 3 (4), 351–360.

Henderson, G., Jin, S., Akin, J., Li, X., Wang, J., Ma, H., He, Y., Zhang, X., Chang, Y., Ge, K., 1995. Distribution of medical insurance in China. Social Science & Medicine 41 (8), 1119–1130.

Liu, X., Mills, A., 1999. Evaluating payment mechanisms: how can we measure unnecessary care? Health Policy and Planning 14 (4), 409–413.

Liu, G.G., Zhao, Z., 2006. Urban employee health insurance reform and the impact on out-of-pocket payment in China. International Journal of Health Planning and Management 21 (3), 211–228.

Liu, G., Cai, R., et al., 2001. China's urban health insurance experiment in Zhenjiang: cost and utilization analyses. In: Hu, T., Hsieh, C. (Eds.), Economics of Health Care Reform in Pacific Rim. Edward Elgar, England, pp. 143–158.

Liu, G.G., et al., 2003. Urbanization and health care in rural China. Contemporary Economic Policy 21 (1), 11–24.

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

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

Mullahy, J., 1997. Instrumental-variable estimation of count data models: applications to models of cigarette smoking behavior. Review of Economics and Statistics 79 (4), 586–593.

Mullahy, J., 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. Journal of Health Economics 17 (3), 247–281.

O'Donnell, O., van Doorslaer, E., Rannan-Eliya, R.P., Somanathan, A., Garg, C.G., Hanvoravongchai, P., Huq, M.N., Karan, A., Leung, G.M., Tin, K., Vasavid, C., 2005. Explaining the incidence of catastrophic expenditures on health care: Comparative evidence from Asia. EQUITAP Working Paper #5. Eramsus University, Rotterdam and IPS, Colombo.

Preker, A.S., Carrin, G., 2003. Health Financing for Poor People: Resource Mobilization and Risk Sharing. World Bank, Washington, DC.

Rubin, R.M., Koelln, K., 1993. Determinants of household out-of-pocket health expenditures. Social Science Quarterly 74 (4), 721–735.

Sepehri, A., Sarma, S., Simpson, W., 2006. Does non-profit health insurance reduce financial burden? Evidence from the Vietnam Living Standards Survey Panel. Journal of Health Economics 15 (6), 603–616.

Shen, Y.C., McFeeters, J., 2006. Out-of-pocket health spending between low-and higher-income populations: who is at risk of having high expenses and high burdens? Medical Care 44 (3), 200–209.

Van Doorslaer, E., O'Donnell, O., Rannan-Eliya, R.P., Somanathan, A., Adhikari, S.R., Garg, C.C., Harbianto, D., Herrin, A.N., Huq, M.N., Ibragimova, S., Karan, A., Lee, T.J., Leung, G.M., Lu, J.F., Ng, C.W., Pande, B.R., Racelis, R., Tao, S., Tin, K., Tisayaticom, K., Trisnantoro, L., Vasavid, C., Zhao, Y., 2007. Catastrophic payments for health care in Asia. Journal of Health Economics 16 (11), 1159–1184.

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

Waters, H.R., Anderson, G.F., Mays, J., 2004. Measuring financial protection in health in the United States. Health Policy 69 (3), 339–349.

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

Zhang, J., 2007. The Impact of a Diagnosis-Related Group-Based Prospective Payment Experiment: The Experience of Shanghai. Graduate School of Economics, Osaka University, Mimeo, Osaka.