

Do Health Sector Reforms Have Their Intended Impacts? The World Bank's Health VIII Project in Gansu Province, China

by

Adam Wagstaff^a and Shengchao Yu^b

^a Development Research Group and East Asia Region, The World Bank, Washington DC, USA

^b World Bank Office Beijing, China

Abstract

The literature contains very few impact evaluations of health sector reforms, especially those involving broad and simultaneous changes on both the demand and supply sides of the sector. This paper reports the results of a World Bank-funded health sector reform project in China known as Health VIII. On the supply-side, the project combined infrastructure investments (especially at the township level) with improved planning and management, including a referral system between township health centers and county hospitals, and interventions aimed at improving the effectiveness and quality of care, including the introduction of clinical protocols and essential drug lists. On the demand-side, the project sought to resurrect community health insurance, and to introduce a safety net for the very poor to provide them with financial assistance with their health care expenses. The evaluation reported here concerns just one of the project's seven provinces, namely Gansu, the reason being that no suitable data are available to undertake a rigorous evaluation in all provinces. This paper makes use of a panel dataset collected for quite another purpose but whose timing (just around the time the project started and four years later) and location (covering both project and non-project counties) makes it well suited to the task. The paper compares estimates obtained using a variety of different estimators, including naïve single differences (before and after, and with and without the project), and differences in differences, adjusting for heterogeneity through both regression and matching methods. The results suggest that it makes a difference to the estimated impact of Health VIII which estimator is used, with the naïve single differences producing often markedly different estimates from the preferred approach of combining difference-in-differences with matching. The results suggest that Health VIII has been mostly successful in its goals. The preferred estimator suggests that the project reduced illness among children, improved self-assessed health and increased doctor visits among the population in general, and reduced the incidence of catastrophic health spending, defined as annual spending in excess of 10% of annual per capita income. However, the project appears to have increased the development and use of high-level facilities, hastened the demise of the village clinic, and may have reduced immunization rates.

Authors' contact details: Adam Wagstaff, World Bank, 1818 H Street NW, Washington, D.C. 20433, USA. Tel. (202) 473-0566. Fax (202)-522 1153. Email: awagstaff@worldbank.org.

Keywords: Impact evaluation; health sector reform; China.

World Bank Policy Research Working Paper 3743, October 2005

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

Acknowledgements: Wave 1 of the Gansu Survey of Children and Families (GSCF) was funded by The Spencer Foundation Small and Major Grants Programs, and Wave 2 by the Fogarty International Center at the National Institutes for Health and the World Bank Research Committee. We are grateful to the GSCF team for allowing us to use the data, and to Lei (Lydia) Liu for help in assembling the NBS county data.

I. INTRODUCTION

With national governments, donors and the international development community growing ever more anxious to see hard evidence of the impact of public programs, and with health featuring so prominently in development objectives, it is reassuring that some recent impact evaluations in the developing world have involved the health sector. The picture could, however, be much rosier than it is.

One problem is that the methodologies used are not always clearly documented, so that one cannot always be sure how reliable the results are. For example, a recent high-profile collection of 17 studies (Levine 2004), purporting to show considerable impact of public health interventions in terms of premature deaths averted, contains no information whatsoever on the methods used in the studies. Where the methods used in health sector impact evaluations *have* been subjected to critical scrutiny, they have often been found inadequate. For example, a recent review (Kapoor 2002) of impact evaluations conducted by the World Bank's independent Operations Evaluation Department (OED) over the last 25 years found that neither of the two health sector projects that had been evaluated had been done so in a rigorous way. Moreover, among the relatively few health sector impact evaluations known to be rigorous, very few concern health system reforms, being more likely to concern the impacts of inputs in the health production function, or the effects on health outcomes of policy changes outside the health sector.¹ There are exceptions. Newman et al. (2002) report impacts on child mortality of health facility infrastructure investments. Saadah et al. (2001) report the impacts on utilization of

¹ Examples of the former include the paper by Jalan and Ravallion (2003) which looks at the effects of piped water on diarrheal disease among children, and the paper by Miguel and Kremer (2004) which looks at the effects of deworming treatment. Examples of the latter include Case's (2002) study of the effect of an old age pension program on inter alia the health of members of the pensioner's household, and the study by Galiani et al. (2005) of the effect on child mortality of the privatization of water services.

Indonesia's health card introduced after the economic crisis of the late 1990s. Gertler (2004) reports the effects on health outcomes of a conditional cash transfer program that required mothers to take their children for regular health checks to receive the cash supplement. And Wagstaff and Pradhan (2005) examine the effects on health utilization and health outcomes of a social health insurance program. Such studies are, however, relatively few in number. Furthermore, all concern a relatively small policy adjustment—none looks at a system reform of the type where several changes are introduced together, possibly operating on the demand and supply sides simultaneously.^{2,3} And yet much of what national governments and donors do in the health sector involves making broad changes to health systems. Over the period 1995-2005, for example, 40% of the World Bank's health sector lending was classified as being directed at “[improving] health system performance”.⁴

This paper reports the results of an impact evaluation of a World Bank-financed health sector reform project in China, known officially as the World Bank China Basic Health Service Project but more often referred to simply as ‘Health VIII’. The evaluation in the paper is partial in that it covers just one of the seven provinces where the project operated, namely Gansu. The reason for the focus on Gansu is the lack of suitable data to undertake a full-scale impact evaluation of Health VIII. The project collected baseline household data only in project counties—in fact, in only 28 of Health VIII's 71 counties.⁵ Even now, seven years after the start

² As of April 27, 2005, the World Bank's impact evaluation database listed 41 impact evaluations of relevance to the health, nutrition and population sector. Not one of these fell into the category “Health Reform and Financing”. The database is available online at <http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm>.

³ See Ravallion (2005) for a recent review of impact evaluations, including those in the health sector.

⁴ In the richer IBRD countries, the share rises to 50%. The classification system is not, it has to be acknowledged, watertight. Some projects concerning communicable disease control are listed under this subheading, though this may be due to the fact they contain components aimed at health system strengthening. It is also possible that some projects that involve health system reform get classified under non-health heads in the bank's system, such as private sector development. My thanks to Lucia Kossarova for providing the breakdown of Bank lending.

⁵ Health VIII began initially in October 1998 in 28 counties, and was subsequently extended to a further 43 counties in late 1999. The official project baseline covers only the initial 28 counties. A further 25 (part B) counties were added to the project even later. As in the government's report of Health VIII, the present paper when referring to Health VIII is referring to the 71 counties joining the project in 1998 and 1999.

of the project, no follow-up household data have been collected; there are, in fact, no plans currently to collect any. The (panel) household data used in this paper—from the Gansu Survey of Children and Families (GSCF)—were collected by researchers quite independently of the Health VIII project. Serendipitously, the GSCF covers both project and non-project counties, before and after implementation of the Health VIII project. The case of Gansu is not without interest. Gansu is home to 26 million people—larger in population terms than most countries⁶—and is China’s second poorest province. Given the emphasis in Health VIII’s objectives of improving health among China’s rural poor, knowing whether Health VIII worked in Gansu is not inconsequential. And while one would like to know something about the impact of the project in the other six project provinces, the fact is that the required data are not—and never will be—available to undertake a rigorous full-scale impact evaluation of Health VIII. In the circumstances, knowing something about the project’s impact in one province seems a distinct improvement on the alternative.

Health VIII marked a break with the past in terms of the World Bank’s support to China’s health sector, and provides an opportunity to study the impact of a health sector reform that is broader than those that have been the subject of impact evaluations to date. In contrast to previous Bank health projects in China that had focused on specific diseases such as Tuberculosis or on specific groups such as women and children, Health VIII was a system-wide project.⁷ Its goals were to raise the quality, affordability and utilization of health services, especially among the poor, and to promote ‘financial protection’ by strengthening risk-pooling arrangements. On the supply-side, the project combined infrastructure investments (especially at township level) with improved planning and management, including a referral system between

⁶ Only 37 countries in the world have a larger population than the province of Gansu.

⁷ Health VIII was selected by the Bank as one of its ten best projects across all sectors in 1998. Details of the project are available in English in Ministry of Health (undated).

township health centers and county hospitals, and interventions aimed at improving the effectiveness and quality of care, including the introduction of clinical protocols and essential drug lists. On the demand-side, the project sought to resurrect community health insurance, and to introduce a safety net for the very poor to provide them with financial assistance with their health care expenses.

In this paper, we use household data to assess whether Health VIII impacted favorably on health service utilization, health outcomes, and out-of-pocket health spending. We also use village data to assess the impact of Health VIII on the availability and perceived quality of health services at village level, the perceived quality of care at township level, and immunization rates. The paper is organized as follows: section II outlines the methods used; section III outlines the data; section IV presents the results; and the final section contains a discussion and our conclusions.

II. METHODS

We compare the results obtained using a number of different estimators. The simplest are based on single differences. The first is the *before-and-after difference*, which is the mean change in outcome before and after the project within the project counties. This is the estimator that was presumably in the minds of those responsible for the project's evaluation when they decided to field the Health VIII baseline survey only in the project counties, and was the estimator used in the evaluation of the Bank's project in China on Schistosomiasis control (Xianyi, Liying et al. 2005). Its obvious limitation is that it assumes that no changes occurred in areas not covered by the project, a limitation that is evident in the evaluation of the Bank's China TB control project (China_Tuberculosis_Control_Collaboration 2004): pulmonary TB *fell*

considerably in non-project counties, so that making a counterfactual assumption of no change in TB in non-project counties would have led to a substantial overestimate of the impact of the project. As will be seen below, reliance on the before-and-after difference estimator in the Health VIII context would also lead to some misleading conclusions.

The second single difference estimator we use is the *with-and-without difference*, which compares the mean outcomes (after the project) between those subjected to Health VIII ('the treated') and those not subjected to the project ('the untreated'). Let y_{it} be the outcome of interest for individual i at time t ($t=1,2$), and T_i equal one if individual i lives in a project county and zero otherwise. Then the *with-and-without difference* estimator can be implemented by means of the convenient regression:

$$(1) \quad y_{it} = \alpha + \gamma T_i + \varepsilon_{it} \quad t=2,$$

which is estimated using data from the second (post-treatment) period only. One obvious drawback of this is that it attributes any difference to the intervention, whereas differences may be due to other factors. An obvious way to try to take such factors into account is to add a set of covariates, x_{kit} , to eqn (1), to get:

$$(2) \quad y_{it} = \alpha + \gamma T_i + \sum_k \phi_k x_{kit} + \varepsilon_{it} \quad t=2.$$

If the error term is uncorrelated with the Health VIII treatment indicator, T_i , the coefficient γ estimates the average treatment effect (cf. e.g. Wooldridge 2002).⁸

The problem is that this assumption is unlikely to be satisfied. It would not be warranted in the present context if people in the Health VIII counties share common characteristics that (a) influence outcomes and (b) are either observable but omitted from x_k or unobservable. As Health VIII was not placed randomly in Gansu's counties (this issue is discussed further below), this is a distinct possibility. One way to take into account selection on observables is to use a *control function estimator* (cf. e.g. Wooldridge 2002). This involves including, alongside the x_k in eqn (2), transformations of these variables interacted with T . In what follows we follow

⁸ All the results in the paper relate to the average treatment effect which is a weighted average of the effect of treatment on the treated and the effect of treatment on the untreated (cf. e.g. Wooldridge 2002).

Wooldridge's (2002) approach and include lagged values of the x_k as well as de-meanned values of the lagged x_k interacted with the Health VIII treatment variable, T :

$$(3) \quad y_{it} = \alpha + \gamma T_i + \sum_k \phi_k x_{kit-1} + \sum_k \mathcal{G}_k (x_{kit-1} - \bar{x}_{kt-1}) \cdot T_i + \varepsilon_{it} \quad t=2.$$

This approach is useful in the present study, because although we observe at two dates variables that are plausible candidates for the x_k , there are some outcome variables that we observe only in the second wave of the GSFC—i.e. *after* the Health VIII project started.

For most outcome variables, we have longitudinal data from before and after the implementation of Health VIII. For these indicators, we can employ the *double-difference* or *difference-in-differences* estimator. This compares the mean before-and-after change among people living in the project areas with the mean before-and-after change among people living in non-project areas. Let P_t equal one if period t is after the project has been implemented. Then the double-difference estimator can be implemented by the convenient regression (cf. e.g. Cameron and Trivedi 2005):

$$(4) \quad y_{it} = \alpha + \beta P_t + \gamma T_i + \delta P_t \cdot T_i + \varepsilon_{it} \quad t=1,2$$

where the interaction term $P_t T_i$ equals one for the treated individuals in the post-intervention period, and the coefficient δ is the difference-in-difference estimate. Or the double-difference estimator can be implemented by regressing the change in outcome over time on a treatment dummy:

$$(5) \quad \Delta y_{it} = \beta + \delta T_i + \Delta \varepsilon_{it} \quad t=1,2,$$

where Δ is the difference operator. The double-difference estimator sweeps out the effects of time-invariant influences on outcomes, both observed and unobserved, and in effect nets out any changes that could be considered likely to have occurred anyway. One can also add covariates to eqns (3) and (4) (cf. e.g. Cameron and Trivedi 2005), to get, for example:

$$(6) \quad y_{it} = \alpha + \beta P_t + \gamma T_i + \delta P_t \cdot T_i + \sum_k \phi_k x_{kit} + \varepsilon_{it},$$

which we refer to below as the double-difference estimator with covariates.

Combining differencing with covariates is just one way of controlling for heterogeneity. An alternative—and cleaner and less restrictive—approach is to combine *differencing with matching* (cf. e.g. Heckman, Ichimura et al. 1997; Imbens 2004; Ravallion 2005). In contrast to the regression approach, no functional form for the outcome variable need be assumed. We combine matching with double differencing but also with single differencing. For outcomes that we observe only in wave 2, computing matched single (i.e. post-intervention) differences is the

best we can do in a matching approach. For outcomes where we observe outcomes before and after the start of Health VIII, the matched single differences will serve as some sort of check on how much faith we should have in the matched single differences for those outcomes where we have only post-intervention data.

The idea behind matching is to compare individuals in project areas with similar individuals in similar non-project areas. So, for example, in computing the differences underlying the average effect of treatment on the treated, we use only *matched* untreated individuals, not *all* untreated individuals. We perform the matching at two levels—the county and the individual. We first run a probit regression across all 76 of Gansu’s counties to predict the probability of the county being a Health VIII project county. We then match individuals on the Health VIII county propensity score as well as pre-intervention values of household- and individual-level variables, using the Mahanobolis metric to measure the closeness of matches across these several dimensions (cf. e.g. Adadie, Drukker et al. 2004). This permits a tradeoff between county characteristics on the one hand and individual and household characteristics on the other: one might end up choosing as a comparison individual someone from a county that has a relatively low probability of being a Health VIII county but who in terms of individual and household characteristics is so close to the treated individual that choosing as a comparison someone from another county would lead to a less good match. In the event, as will be seen, the matching procedure results in us choosing as comparisons people in a handful of non-project counties that have propensity scores that are very close to those of the Health VIII counties.

III. DATA

The Health VIII project operates in 10 of Gansu's 76 counties. Three of these (Dingxi, Kang and Wudu), have been surveyed in the GSFC. One (Wudu) was one of the original 28 first-wave Health VIII counties: in these counties, the project began in October 1998. The other two counties (Dingxi and Kang) joined Health VIII in late 1999. The first wave of GSFC data were collected in June 2000, somewhat after the start of Health VIII, but given its complexity and scope, it seems unlikely that much—if any—impact will have been felt by the time of the first wave of the GSFC, especially in Dingxi and Kang. In addition to collecting data in the three Health VIII counties, the GSFC also collected data from households in 17 of Gansu's 66 non-project counties. The same households were then revisited in mid-2004, and a second wave of data were collected.

The GSFC panel contains 1,116 individuals (186 households) living in the three Health VIII counties, and 6,465 individuals (1,148 households) living in the 17 non-project counties. The household questionnaires contain key information for all household members on health outcomes, service utilization and household health spending. In addition, we have data on health and health service utilization from the second wave on 186 GSFC 'target' children. In addition to these datasets, we have some village-level health data from questionnaires administered to the leaders of the 100 GSFC villages. These provide a useful complement to the household data.

The individual-level outcomes for all household members that we examine include health outcomes, doctor visits, and out-of-pocket health spending on outpatient care and medicines. These are detailed in Table 1. The self-assessed health variable is of the type used in a variety of surveys in industrialized countries, as well as in several developing country surveys such as the Indonesia Family Life Survey and the China Health and Nutrition Survey. In industrialized

countries, at least, it has been found to be a good predictor of mortality and the onset of disability (cf. Idler and Benyamini 1997). The chronic illness variable was defined differently in the two waves, which makes its use somewhat problematic in the double differences. In addition to looking at the impact of Health VIII on mean or expected out-of-pocket health spending, we also examine its impact on the probability of the individual having catastrophic health spending, defined here as spending exceeding 10% of annual per capita income (cf. e.g. Wagstaff and van Doorslaer 2003).

Two of the health outcome indicators moved in opposite directions to the other two: the self-assessed health and disability indicators suggested a worsening in health outcomes, while the chronic illness variable and the number of sickness days suggest the opposite. The dramatic reduction in chronic illness probably reflects the change in the way the question was posed, with a list being used in wave 2 but not in wave 1. The fall in the number of doctor visits and the increase in expenses on doctor visits are in line with trends reported from other China surveys. The significant decline in drug outlays is rather surprising, and may reflect a different allocation between the two expense categories over the two waves. We explored the sensitivity of the results to combining the two categories but found little difference. The catastrophic spending indicator points to a reduction in very large expenses between the two waves.

The outcomes variables for the target child are listed in Table 2. These are available for the second wave only. We explore the impact of Health VIII on the probability of the child being taken to the doctor in the previous 12 months, with and without controlling for illness. In the matching approach, to control for illness, we match exactly on illness. We explore the impact on type of provider used only among those reporting some visits—in the matching approach, we do this by matching exactly on the dummy doctor visits variable.

The household- and individual-level covariates that we use in the regression and matching analyses (Table 3) suggest rising incomes, but also point to a sharp rise in insurance coverage. Those who have insurance in the second wave typically have school insurance, the annual premium for which is recorded in the survey, and is very low (RMB 30)—well below the premium paid for other types of insurance, likely reflecting a limited set of risks covered, presumably the relatively inexpensive childhood illnesses and preventive activities.

In common with villages elsewhere in China, GSFC villages have tended to lose their clinic (Table 4). Where clinics have survived, they have tended to increase their staff numbers, though not significantly so in the case of midwives. Villages have also shifted their preferences somewhat towards lower-level providers, but the change is not marked. According to data from the village leader interviews, there has been a significant drop in immunization coverage in the GSFC villages in the period in question.⁹ Data on perceptions by village leaders of the quality of medical care are available only for the second wave. Interestingly, the quality of care in village clinics is rated more highly than that in township health centers.

For the matching we also require a county-level variable indicating the probability of the county being selected as a Health VIII county. The criteria for selecting a county for inclusion in Health VIII are spelt out in the project documentation. Poor counties—as reflected in whether they are nationally or provincially designated poverty counties—are preferred, *ceteris paribus*. All Health VIII counties in Gansu are, in fact, poverty counties. Counties with poor health outcomes, as reflected in high child mortality, are also preferred. How this is operationalized is something of a mystery, as reliable county-level data on child mortality do not exist, or at least

⁹ The recent falls in immunization coverage in poorer parts of rural China were commented on in the 2004 International Review of China's Expanded Program on Immunizations (EPI) conducted by the World Health Organization, the United Nation's Children Fund, the Global Alliance for Vaccine Initiative, the Japan International Cooperation Agency and the U.S. Centers for Disease Control.

are not in the public domain. A further consideration is the county's fiscal position. In China, it is the central government that borrows from the World Bank. Beijing then on-lends to provinces, charging a fee in the process, and the province then on-lends to the county, again charging a fee for its services. Counties with very limited resources will not be able to repay the provincial government, and are therefore less likely *ceteris paribus* to end up with a World Bank project in health or any other sector. The final consideration is the county's capacity to implement the project, with counties with high capacity being preferred. This increases the likelihood of the project achieving its objectives, but seems likely to tilt the scales against poorer counties.

Table 5 reports the results of two county-level probit regressions (data refer to 2000). The second is included because the very high correlation among the covariates makes it hard to detect their independent effects—basically, per capita income is either a cause or consequence of most of the indicators. The results confirm that, within poverty counties, richer counties have a higher probability of being a Health VIII county, and suggest strongly that counties with more health sector capacity, as proxied by the number of hospital beds, also have a higher probability of being selected for inclusion. Health VIII is clearly not reaching the poorest of the poor. From the probit equation, a propensity score is computed for each county, using the full model rather than the more parsimonious model.

Table 1: Outcome variables—individual-level from household questionnaire

Variable	Definition	Comment	Mean Wave 1	Mean Wave 2	t-test
Overall assessment of health	Very poor (1), poor, average, good, very good (5).	Assessment is by one respondent for all household members.	4.136	3.971	-11.94
Chronic illness	No (0), yes (1). See comment.	In wave 1, respondents were asked whether in the past year they had suffered from any chronic disease. In wave 2, they were asked whether they suffered from any of the following: cancer, heart disease, diabetes, hepatitis or other. If the answer was none, we coded them as not having a chronic illness in wave 2.	0.064	0.037	-8.32
Disability	Classified as disabled if respondent says yes to any of the following types of disability read out to the respondent by the interviewer: deaf or mute; blind; bodily disability; mental illness; retarded; or 'other'.		0.125	0.158	2.62
Sickness days	No. days in past 3 months when sickness precluded respondent from carrying out his or her daily activities, such as work or school.		1.637	1.202	-3.18
Doctor visits	No. visits in the last year to a doctor.	In Wave 1 respondents were initially asked whether they had stayed in a hospital, and then asked whether they had seen a doctor. Presumably the hospital question in wave 1 refers to inpatient episodes and the doctor visit question in both waves to outpatient visits.	1.867	1.049	-12.67
Doctor visit expenses	Amount spent in last 12 months by the household on behalf of respondent for doctor visits.	In wave 1 each household member was asked how much had been spent on medical advice. Whether this could have included expenses associated with inpatient care is unclear. In wave 2, each household member was asked how much had been spent on seeing a doctor. Explicitly listed as examples but not separately itemized were the registration fee, diagnosis fee, examination fee, and cost of medicines obtained from doctor.	77.643	156.044	4.28
Drug expenses	Amount spent by the household on behalf of the respondent in last 12 months purchasing medicines.	It is not explicit but presumably the case that medicines obtained during doctor visits are included under 'doctor visit expenses' and not under 'drug expenses'.	66.726	55.486	-2.34

Table 1 (contin): Outcome variables—individual-level from household questionnaire

Variable	Definition	Comment	Mean Wave 1	Mean Wave 2	t-test
Catastrophic medical expenses	Equals 1 if annual expenses associated with doctor visits and drugs exceeded 10% of household per capita income.		0.189	0.141	-8.07

Table 2: Outcome variables—GSFC target children

Variable	Definition	Comment	Mean Wave 2
Ill in last year	No (0), yes (1).	Mother is asked whether during the last 12 months the child has been diagnosed by a doctor with the following: anemia, asthma or other chronic respiratory diseases, TB, pneumonia or other acute respiratory diseases, cold, injury due to accidents, diarrhea, eye illness, or parasite disease. The 'ill' variable takes a value of 1 if any of these conditions is recorded.	0.585
Taken to doctor	No (0), yes (1).	Mother was asked whether child had been taken during the last 12 months to a doctor to receive medical care. Question is asked irrespective of whether child has been ill in last 12 months.	0.408
Highest level of provider used during last 12 months	0=none; 1=village clinic; 2=township health center; 3=county hospital	Mean in last column is among users only.	1.717

Table 3: Household and individual covariates for outcome equations and matching

Variable	Definition	Note	Mean Wave 1	Mean Wave 2	t-test
Sex	Male = 1		0.515	0.515	
Age	Age in years		25.295	29.269	
Health insurance	1 if the household member is covered, 0 otherwise	In wave 1, each household member was asked how their health care was paid for: self-paid, cooperative medical scheme (CMS), health insurance, free at public expense (includes GIS), or other. Household member is classified as insured if he or she has CMS, health insurance, or free at public expense. In wave 2 the household as a whole was asked whether any family member had health insurance including CMS. A follow-up question is then asked of <i>each</i> member who has insurance establishing the type of insurance that they have (employer, rural CMS and rural health insurance, private insurance, school health insurance, or other). Household member is classified as insured on the basis of answers to these two questions.	0.003	0.182	39.80
Years of schooling	Years spent by respondent in education		13.561	16.771	50.48
Household per capita income	Income from agriculture, livestock, wages and self-employment.	Wage income includes bonuses, subsidies, and the value of in-kind payments.	1844.291	3351.850	8.00

Table 4: Village-level variables

Variable	Definition	Mean Wave 1	Mean Wave 2	t-test
Clinic exists in village	=1 if a clinic exists in the village	0.95	0.89	1.60
# doctors in nearest village clinic	# doctors working in village clinic or nearest village clinic if none in village	2.74	3.69	2.37
# nurses in nearest village clinic	# nurses working in village clinic or nearest village clinic if none in village	0.74	2.53	5.64
# midwives in nearest village clinic	# midwives working in village clinic or nearest village clinic if none in village	1.95	2.08	0.13
Provider preference by villagers	When villagers need to see a doctor, where do they usually go? 1=village doctor; 2=township hospital; 3=pharmacy nearby; 4=county hospital	1.82	1.60	1.60
Immunization rate in village	Immunization rate for children in the village	97.90	94.69	2.42
Quality of village clinic	What is the village leader's rating of the quality of medical care in the village clinic or in the clinic of the village nearby? 1=bad; 4=very good		2.40	
Quality of township health center	What is the village leader's rating of the quality of medical care in the township hospital? 1=bad; 4=very good		2.16	

Table 5: County-level probit equation used to predict placement of Health VIII

Variable	Mean	Coefficient	z-statistics	Coefficient	z-statistic
Fraction of the population classified as rural	0.83	10.22119	1.04	10.24852	1.97
GDP per capita	2905.49	0.00055	1.60	0.00035	1.71
Per capita local government revenue	201.61	-0.00094	-0.34		
Per capita local government spending	524.78	-0.00032	-0.19		
Per capita completed investment in construction	625.89	0.00047	0.70		
Per capita accumulated savings	2145.86	-0.00045	-0.77		
Telephone connections per 10,000 population	480.10	-0.00093	-0.43		
Hospital beds per 10,000 population	19.06	0.07831	1.40	0.04856	1.86
Social welfare institution beds per 10,000 population	2.52	-0.25324	-1.70		
Primary school enrollment per 10,000 population	1194.45	-0.00179	-0.94		
Secondary school enrollment per 10,000 population	438.42	0.00112	0.52		
Pseudo R ²		0.3445		0.2104	

Note: Data are from National Bureau of Statistics county database and refer to 2000. Sample includes all Gansu counties, not just those in GSFC.

IV. RESULTS

This section presents the individual-level and village-level estimates of the impact of Health VIII on key outcomes. A discussion of the results—whether they are plausible and possible reasons for them—is postponed until section V.

Table 6 presents the impact results for the individual-level variables. One important finding is worth highlighting immediately—the results for several outcomes vary considerably depending on which method is used. For example, if one focuses on the before-and-after change in self-assessed health, Health VIII appears to have worsened health, albeit not significantly. But as is clear from the the results obtained through double-differences without covariates added, the non-project counties experienced a much larger deterioration in self-assessed health than did the Health VIII counties. Indeed, the estimators other than the before-and-after difference suggest that self-assessed health *improved* as a result of Health VIII. The before-and-after estimator also gives misleading results for other outcomes. It suggests, for example, that Health VIII reduced the number of doctor visits—by as much as 1.4 visits per person per year. The other estimators suggest either that Health VIII reduced the number of visits but by much less, or in the case of the double-difference *increased* the number of visits. But it is not just the before-and-after difference that gives misleading results. The other single-difference—the post-intervention difference between the Health VIII and non-project counties—is also misleading in several cases.

One key influence on many results is whether matching is used rather than adding covariates to a regression equation. The reason for this is that the matching approach restricts comparisons to a subset of GSFC counties that are similar to Health VIII counties. Table 7 shows

for each Health VIII county the propensity score, the number of sampled individuals, and how they are matched in computing the treatment effect for the treated across non-project counties. Also shown are the propensity scores of the non-project counties. (The counties not used as matches are not listed.) So, for example, most of sampled individuals in Wudu county are matched with individuals in Qin'an county whose propensity score comes closest to that of Wudu. However, a few are matched to people in Yongjing county, because although Yongjing's propensity score is further from Wudu's than that of Qin'an, the individuals found as matches in Yongjing are on balance more similar than the best matches that can be found in the GSFC in Qin'an county, taking into account the county propensity score, the individual characteristics *and* the household characteristics. Overall, Qin'an and Yongjing provide the bulk of matches for the three Health VIII counties. Given this, it is unsurprising that the results obtained using the matching method are often quite different from those obtained using other methods where people in *all* non-project counties are used in computing the relevant difference.

Focusing on the preferred matched double-difference estimates, Table 6 suggests that Health VIII significantly *increased* the number of doctor visits, but did *not* significantly reduce expected health spending, chronic illness or sickness days. Like other estimators, the matched difference-in-differences estimator also suggests that Health VIII led to a statistically significant improvement in self-assessed health and to a reduction in the incidence of catastrophic health spending. The results for the target child¹⁰ on health status are consistent with these results (Table 8), suggesting that Health VIII reduced the probability of illness. However, the results also suggest that the project significantly reduced the probability of a doctor visit, although not conditional on being ill. The target child results also suggest that Health VIII encouraged

¹⁰ The covariates and matching variables used are the same as those used in Table 6 but are household averages. So, for example, it is not the child's own years of education that is used but rather the average years of all household members.

families who did take their child to a provider to seek a higher-level provider than would otherwise have been the case.

This is a result that comes through too in the village-level results, though the impact in the case of the village data is not significant (Table 9).¹¹ What the village level data *do* suggest is that irrespective of the estimator used, Health VIII hastened the collapse of the village clinic. Conditional on there being a clinic, there is also a hint that Health VIII also pushes down the number of staff working in village clinics, though the matched double-difference results are not significant for any category of staff.

One further result that emerges from Table 7—and a worrying one if true—is that Health VIII may have had a negative impact on immunization coverage in Gansu.¹² The results vary according to the estimator used, but the fact that the preferred matched double-difference estimator produces a statistically significant negative average treatment effect is worrying. The final set of village-level results concern the perceived quality of medical care in village clinics and township hospitals. The evidence is limited to the post-treatment comparisons, because data are available only for the second wave, but we take some comfort from the fact that the matched post-treatment and matched double-difference estimates are not too dissimilar for those village-level outcomes where data are available for both waves. The post-treatment matched difference results suggest that Health VIII improved the quality of care in both village clinics and township

¹¹ The matching in the case of the village-level indicators was done using the county propensity score, the poverty county dummy, and village average per capita income from the household data.

¹² This is one area where generalizing to other Health VIII provinces would be dangerous. Within the broad scope of Health VIII, different counties gave priority to different packages of medical interventions. None of the project counties in Gansu selected immunization as their priority area. It may well be that the apparent negative effects of Health VIII in Gansu were avoided in counties giving priority to immunization. These were, however, a minority of Health VIII counties. Only 20 out of 70 had selected immunization as a priority area, each county being able to select more than one area. So, the result is relevant to the majority of Health VIII counties, and even if it were not true of the counties where immunization was selected as a priority, it serves as a warning that unless prioritized, immunization could suffer from the type of reform package introduced by Health VIII.

hospitals, but not significantly so. There is a hint that the impact was larger in the case of the township hospital.

Table 6: Results for individual-level data

	Before-and-after difference—Health VIII counties only	Post-treatment difference between Health VIII and non-project counties. Eqn (1)	Control function estimator. Eqn (3)	Difference-in-differences. No covariates. Eqn (4)	Difference-in-differences. Covariates added. Eqn (6)	Difference-in-differences. Covariates & county fixed effects added. Eqn (6)	Difference-in-differences with nearest neighbor matching
Health	-0.022 (0.37)	0.203 (3.34)	0.231 (3.52)	0.168 (2.55)	0.171 (2.58)	0.169 (2.57)	0.191 (2.40)
Chronic illness	-0.056 (4.46)	0.024 (2.51)	0.011 (1.28)	-0.033 (2.47)	-0.033 (2.49)	-0.033 (2.48)	0.009 (0.45)
Disability	0.023 (0.48)	0.044 (1.03)	-0.009 (0.23)	-0.012 (0.23)	-0.013 (0.25)	-0.012 (0.23)	0.111 (1.61)
Sick days	-1.536 (3.39)	0.667 (1.86)	0.061 (0.23)	-1.291 (2.70)	-1.272 (2.66)	-1.270 (2.66)	0.487 (0.47)
Doctor visits	-1.405 (8.25)	-0.669 (7.52)	-0.743 (9.74)	-0.688 (3.47)	-0.665 (3.35)	-0.665 (3.35)	0.770 (2.03)
Doctor visit expenses	16.567 (0.91)	-97.529 (3.91)	-99.192 (4.24)	-72.507 (2.56)	-71.594 (2.52)	-71.434 (2.52)	-12.418 (0.21)
Medicine expenses	-59.884 (3.87)	-6.884 (0.70)	-12.148 (1.48)	-57.041 (3.48)	-55.459 (3.39)	-55.101 (3.38)	-5.558 (0.17)
Catastrophic expenses	-0.144 (7.78)	-0.029 (2.69)	-0.073 (6.20)	-0.112 (5.45)	-0.110 (5.48)	-0.109 (5.45)	-0.142 (4.32)

Notes: absolute values of t-statistics or z-statistics in parentheses. Sample includes 1116 individuals living in project counties, and 6465 individuals living in non-project counties.

Table 7: How people living in Health VIII counties are matched to people living in non-project counties

Health VIII counties and propensity score	Non-project counties and propensity score						<i>Total</i>
	Tongwei 0.0375	Tianzhu 0.0000	Qin'an 0.1268	Yongjing 0.1153	Yuzhong 0.0546	Yongdeng 0.0774	
Dingxi 0.9997			389	9		1	399
Wudu 0.2592	2		433	59			494
Kang 0.1048	13	6	73	117	13		222
<i>Total</i>	<i>15</i>	<i>6</i>	<i>895</i>	<i>185</i>	<i>13</i>	<i>1</i>	<i>1,115</i>

Table 8: Results for target child

	Post-treatment difference	Control function	Post-treatment difference with matching
Ill during last year	-0.062 (1.57)	-0.040 (0.76)	-0.104 (1.67)
Child taken to doctor during last year	-0.091 (2.32)	-0.110 (2.27)	-0.153 (2.51)
Child taken to doctor during last year conditional on child being ill		-0.098 (2.36)	-0.046 (0.84)
Provider choice conditional on child being taken to a provider	-0.128 (1.77)	0.076 (1.96)	0.156 (1.97)

Note: Sample includes 186 target children living in project counties, and 1148 target children living in non-project counties.

Table 9: Results for village-level data

	Before-and-after difference— Health VIII counties only	Post-treatment difference between Health VIII and non- project counties. Eqn (1)	Control function estimator. Eqn (3)	Post-treatment difference with nearest neighbor matching	Difference-in- differences. No covariates. Eqn (4)	Difference-in- differences. Covariates added. Eqn (6)	Difference-in- differences with nearest neighbor matching
Clinic exists in village	-0.286 (1.76)	-0.370 (1.86)	-0.310 (1.32)	-0.373 (1.90)	-0.262 (1.84)	-0.262 (1.85)	-0.400 (1.93)
# doctors in nearest village clinic	0.786 (0.70)	-0.719 (1.40)	-2.502 (2.40)	-1.453 (0.85)	-0.191 (0.19)	-0.193 (0.20)	-0.860 (0.51)
# nurses in nearest village clinic	1.714 (2.02)	-0.533 (0.93)	-1.313 (1.28)	0.007 (0.01)	-0.088 (0.11)	-0.088 (0.11)	0.687 (0.55)
# midwives in nearest village clinic	-2.500 (5.35)	-1.754 (1.57)	-2.943 (2.38)	-2.533 (0.63)	-3.058 (2.42)	-3.060 (2.52)	-3.227 (0.77)
Provider preference by villagers	0.143 (0.70)	1.379 (4.48)	0.496 (1.37)	1.713 (3.67)	0.422 (1.72)	0.422 (1.75)	0.760 (1.26)
Immunization rate in village	-6.071 (1.52)	-4.955 (1.20)	17.150 (2.98)	-17.647 (2.24)	-3.323 (0.92)	-3.330 (0.95)	-15.945 (2.14)
Quality of village clinic		0.116 (1.06)	1.520 (4.06)	0.047 (0.15)			
Quality of township health center		0.312 (3.57)	0.692 (4.35)	0.387 (1.22)			

Notes: absolute values of t-statistics or z-statistics in parentheses. Sample includes 14 villages in project counties, and 86 villages in non-project counties.

V. DISCUSSION AND CONCLUSIONS

The methodological conclusion to emerge is not altogether surprising, namely it makes a difference to the estimated impact of Health VIII which estimator is used. Of the different estimators used in the analysis, the double-difference estimator combined with matching is likely to come closest to the truth. Compared to this, the before-and-after difference—the estimator that apparently was in the minds of those setting up the official Health VIII evaluation—does a particularly poor job, especially for the individual-level outcomes. Using this estimator one would have incorrectly concluded that Health VIII had *reduced* self-assessed health and doctor visits, and had also significantly reduced out-of-pocket expenses.

The substantive findings suggest that Health VIII has been partially successful in its goals. The matched double difference estimator suggests that the project improved self-assessed health of the population in general, reduced illness among children, increased doctor visits, and reduced the incidence of catastrophic household health spending. It did *not*, however, significantly affect the mean level of household health spending. The results for the target child and the village suggest that Health VIII increased use of high-level facilities at least among children, hastened the decline of the village clinic, and reduced immunization rates. No significant impacts of Health VIII were found on staffing levels in the village clinics that survived, or on perceived quality of care at either village or township level.

Are these findings plausible? Certainly, the goal of Health VIII was to increase use of health services, and the finding that doctor visits were positively impacted by the project is consistent with this. If doctor visits did indeed improve, it is perfectly plausible that out-of-pocket payments did not fall. Indeed, the fact that they did not *rise*, despite the rise in the number

of visits, points to Health VIII being successful in putting downward pressure on household payments per visit. And if visits increased, and the supply-side investments succeeded in improving the quality of care delivered, it is plausible that health also improved. The individual-level results seem therefore to be consistent with the project's goals, and with the mechanisms put in place to achieve them.

The more surprising and worrying results come from the analysis of the target child and village data. Is it plausible that Health VIII encouraged use of higher-level facilities and accelerated the decline of the village clinic? On the face of it, it would seem unlikely, because the project documentation talks of Health VIII investing at the township *and village* levels. However, in practice the emphasis in the project appears to be on strengthening (public) township facilities rather than (typically private) village providers. And because the Health VIII package of measures aimed at strengthening the township hospital—extra equipment, improved infrastructure, better management, and measures such as clinical protocols aimed at improving the quality of care—implies extra competition for the village doctor, it would not be altogether surprising if in the process Health VIII were to crowd the village doctor out of the market.

Another worrying result is the negative impact of the project on immunization rates. Official immunization data are, it must be said, often unreliable, but it is far from clear why they should have become increasingly unreliable over this time period, and even less clear why the quality of the data should have deteriorated faster in Health VIII counties than in matched non-project counties. In fact, the peculiarities of China's approach to immunization financing and delivery¹³ make it possible that the negative impact on immunization is *not* a spurious result. Vaccinations are delivered by village clinics, township health centers and local public health

¹³ For a useful description of the system and its problems, see the 2004 International Review of China's Expanded Program on Immunizations (EPI) conducted by the World Health Organization, the United Nation's Children Fund, the Global Alliance for Vaccine Initiative, the Japan International Cooperation Agency and the U.S. Centers for Disease Control.

institutes (often called CDCs in English). Village doctors are required to deliver immunization as a condition for their licensure, and receive little or no public finance for this work—they have either to absorb the cost or charge families. CDCs receive a fixed subsidy for immunization and other activities, but they are also free to generate revenues from public health and related activities, and the actual disbursement of the subsidy is not linked to performance vis-à-vis immunization or anything else. China is also unusual by international standards in recommending far more immunization sessions in infancy than WHO recommends (twice as many, in fact). If it is true that Health VIII hastened the demise of the village clinic, it is not inconceivable that, facing the prospect of multiple trips from their village to the township CDC to get their child immunized, fewer families ended up taking their children for vaccination. And, with little incentive to engage in outreach (they earn no extra revenues by increasing immunization coverage but do by selling other services), CDCs are unlikely to have gone looking for them.

All in all, then, a mixed picture emerges of the impact of Health VIII in Gansu province. On the ‘plus’ side, the project appears to have improved health outcomes among adults and children, encouraged use of health services at least among adults, and reduced the incidence of catastrophic household health spending . On the ‘minus’ side, it appears to have encouraged the development of and use of higher-level facilities, and in the process may have impacted negatively on immunization.

References

- Adadie, A., D. Drukker, et al. (2004). "Implementing matching estimators for average treatment effects in Stata." The Stata Journal **4**(3): 290-311.
- Cameron, A. and R. Trivedi (2005). Microeconometrics: Methods and Applications.
- Case, A. (2002). Health, Income and Economic Development. Annual World Bank Conference on Development Economics: 2001/2002. B. Pleskovic and N. Stern. Washington, D.C., World Bank; Oxford and New York: Oxford University Press: v, 269.
- China_Tuberculosis_Control_Collaboration (2004). "The effect of tuberculosis control in China." Lancet **364**(9432): 417-22.
- Galiani, S., P. Gertler, et al. (2005). "Water for life: The impact of the privatization of water services on child mortality." Journal of Political Economy **113**(1): 83-119.
- Gertler, P. (2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." American Economic Review **94**(2): 336-41.
- Heckman, J. J., H. Ichimura, et al. (1997). "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." Review of Economic Studies **64**(4): 605-54.
- Idler, E. and Y. Benyamini (1997). "Self-rated health and mortality: a review of twenty-seven community studies." Journal of Health and Social Behavior **38**(1): 21-37.
- Imbens, G. W. (2004). "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." Review of Economics and Statistics **86**(1): 4-29.
- Jalan, J. and M. Ravallion (2003). "Does Piped Water Reduce Diarrhea for Children in Rural India?" Journal of Econometrics **112**(1): 153-73.
- Kapoor, A. (2002). Review of Impact Evaluation Methodologies Used By the Operations Evaluation Department Over the Last 25 Years. Washington DC, World Bank.
- Levine, R. (2004). Millions Saved. Washington DC, Center for Global Development.
- Miguel, E. and M. Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica **72**(1): 159-217.
- Ministry_of_Health_of_China_(Foreign_Loan_Office) (undated). Basic Health Service Project (Health VIII and Health VIII Support Project): Report of Phase One Review. Beijing, Foreign Loan Office, Ministry of Health of China.
- Newman, J. L., M. Pradhan, et al. (2002). "An impact evaluation of education, health, and water supply investments by the Bolivian Social Investment Fund." World Bank Economic Review **16**(2): 241-74.
- Ravallion, M. (2005). Evaluating Anti-Poverty Programs. Handbook of Agricultural Economics vol. 4. R. Evenson and T. Schultz. Amsterdam, North Holland.
- Saadah, F., M. Pradhan, et al. (2001). The effectiveness of the Health Card as an instrument to ensure access to medical care for the poor during the crisis. Washington DC, World Bank.
- Wagstaff, A. and M. Pradhan (2005). Health Insurance Impacts on Health and Non-Medical Consumption in a Developing Country. Washington DC, World Bank.
- Wagstaff, A. and E. van Doorslaer (2003). "Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998." Health Econ **12**(11): 921-34.
- Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge, Mass., MIT Press.

Xianyi, C., W. Liying, et al. (2005). "Schistosomiasis control in China: the impact of a 10-year World Bank Loan Project (1992-2001)." Bull World Health Organ **83**(1): 43-8.