

Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme



Gary King, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Ryan T Moore, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, Héctor Hernández Llamas

Summary

Background We assessed aspects of Seguro Popular, a programme aimed to deliver health insurance, regular and preventive medical care, medicines, and health facilities to 50 million uninsured Mexicans.

Methods We randomly assigned treatment within 74 matched pairs of health clusters—ie, health facility catchment areas—representing 118 569 households in seven Mexican states, and measured outcomes in a 2005 baseline survey (August, 2005, to September, 2005) and follow-up survey 10 months later (July, 2006, to August, 2006) in 50 pairs ($n=32\,515$). The treatment consisted of encouragement to enrol in a health-insurance programme and upgraded medical facilities. Participant states also received funds to improve health facilities and to provide medications for services in treated clusters. We estimated intention to treat and complier average causal effects non-parametrically.

Findings Intention-to-treat estimates indicated a 23% reduction from baseline in catastrophic expenditures (1·9% points; 95% CI 0·14–3·66). The effect in poor households was 3·0% points (0·46–5·54) and in experimental compliers was 6·5% points (1·65–11·28), 30% and 59% reductions, respectively. The intention-to-treat effect on health spending in poor households was 426 pesos (39–812), and the complier average causal effect was 915 pesos (147–1684). Contrary to expectations and previous observational research, we found no effects on medication spending, health outcomes, or utilisation.

Interpretation Programme resources reached the poor. However, the programme did not show some other effects, possibly due to the short duration of treatment (10 months). Although Seguro Popular seems to be successful at this early stage, further experiments and follow-up studies, with longer assessment periods, are needed to ascertain the long-term effects of the programme.

Funding Mexican Ministry of Health, the National Institute of Public Health of Mexico, and Harvard University Institute for Quantitative Social Science.

Introduction

Health-system reforms often fail to provide the poor with access to quality care.¹ Mexico has attempted to avoid this shortcoming by implementing a new set of health reforms called Seguro Popular,² which aim to provide health coverage to 50 million uninsured Mexicans. The main aim of these health reforms of 2003, is to reduce the prevalence of catastrophic health expenditures by “(providing) social protection in health...”³ Seguro Popular consists of health policies⁴ and priorities,⁵ including entitlements for affiliated families, well-defined benefits packages including coverage for 266 unique health interventions, 312 medicines, increased funds to state health ministries proportional to the number of Seguro Popular-affiliated families, federal funds for personal and non-personal health services, and creation of special federal funds for catastrophic medical expenditures associated with certain diseases.⁶ By linking federal support to medical facility quality, Seguro Popular aims to strengthen an accreditation system for health clinics and hospitals. When rollout is complete, Seguro Popular is intended to increase total health spending in Mexico by a full percentage point of gross domestic product (from 5·6% in 2002).⁷

Seguro Popular sought to build a stronger evidence base by permitting a rigorous, independent, scientific assessment of the programme. Here, we present an experimental assessment of these health reforms. We introduce aspects of a so called stepped wedge experimental design,^{8,9} which uses the phased rollout of the national programme to make randomisation politically feasible and ethical. We also discuss expectations, including the idea that Seguro Popular funds would reach the poor and increase health-care utilisation.

Methods

Experimental design

Using a real-world health policy programme as an experimental treatment offers tremendous advantages but also requires changes in research design and interpretation to anticipate political and implementation challenges.⁸ As discussed by King and colleagues,⁸ “The history of public policy experiments is littered with evaluations torpedoed by politicians appropriately attentive to the short term desires of their constituents, such as those who wind up in control groups without new services or who cannot imagine why a government

Lancet 2009; 373: 1447–54

Published Online

April 8, 2009

DOI:10.1016/S0140-6736(09)60239-7

See Comment pages 1404 and 1406

Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA (Prof G King PhD);

Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA (E Gakidou PhD, N Ravishankar PhD);

Department of Politics, Princeton University, Princeton, NJ, USA (K Imai PhD);

Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA (J Lakin PhD, C Nall MA);

Secretaría de Salud, Mexico, DF, Mexico (M Vargas MPA, M Hernández Ávila MD);

Instituto Nacional de Salud Pública, Cuernavaca, Morelos, Mexico (M M Téllez-Rojo PhD, J E Hernández Ávila MSc, M Hernández Ávila);

Conestadística, Mexico, DF, Mexico (H Hernández Llamas PhD); and Washington University in St Louis, St Louis, MO, USA (R T Moore PhD)

Correspondence to:
Prof Gary King, Institute for Quantitative Social Science, 1737 Cambridge Street, Harvard University, Cambridge, MA 02138, USA
king@harvard.edu

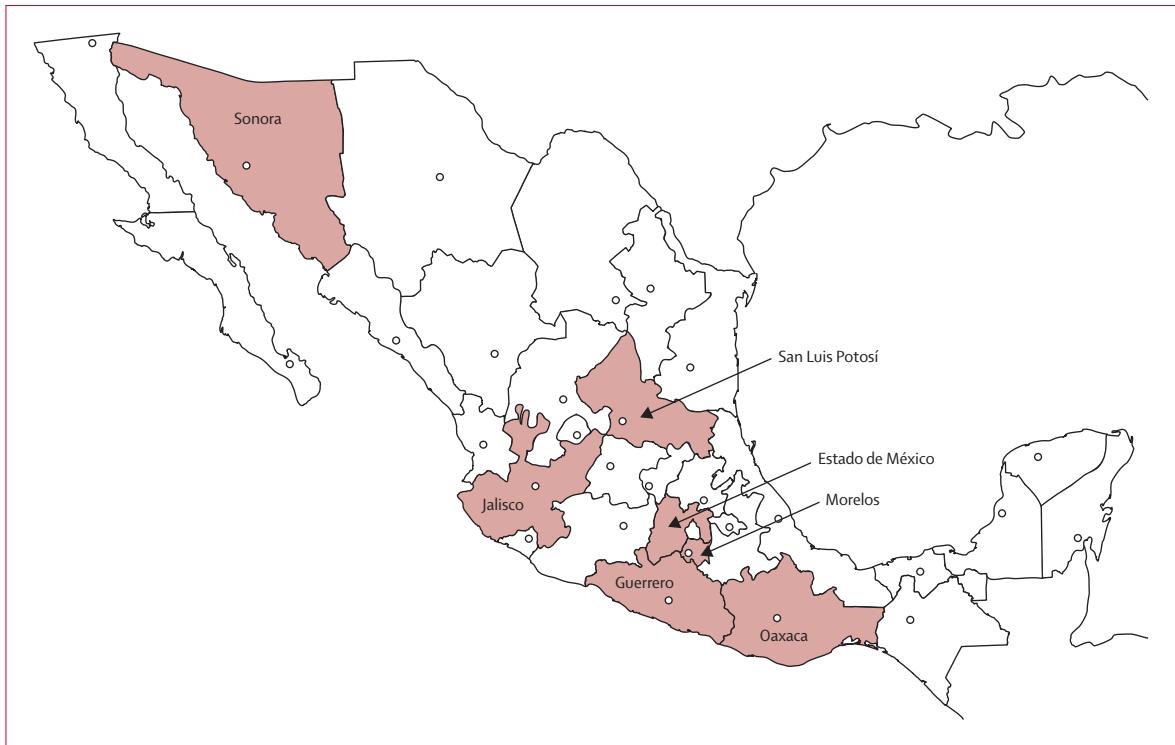


Figure 1: Mexican states participating in the Seguro Popular assessment

would randomly assign citizens to government programs". Our design is adapted to survive these types of interventions.

We developed an experimental design within the phased rollout of Seguro Popular to introduce random assignment. Within selected communities waiting for the programme rollout, we assigned some communities to receive Seguro Popular at least a year earlier than expected, marshalling procedures that would otherwise be based on state implementation plans for clinic accreditation and population coverage. The design is robust to several types of politically-driven experimental protocol violations.⁸

We designed a matched-pair cluster-randomised experiment. First, we defined 12 284 health clusters—contiguous geographical regions of Mexico's 32 states. Each cluster includes a (present or planned) health clinic or hospital and the population in its catchment area. In collaboration with the ministry of health and the Seguro Popular commission, we persuaded 13 Mexican states to join the assessment, including 7078 (5439 rural and 1639 urban) health clusters. We paired health clusters before randomisation, ensuring that matches were as similar as possible to each other for cluster size and various background characteristics.^{10,11} We then negotiated access to 74 cluster pairs in seven states (figure 1), with inclusion based on necessary administrative, political, and other criteria.

We randomly assigned one health cluster from each pair to be the treatment cluster, which meant a campaign to persuade every family to enrol in Seguro Popular, and procedures initiated by states to improve health facilities and increase medical personnel and drug supplies to implement the programme effectively. Poor families in treatment clusters already participating in the pre-existing Oportunidades anti-poverty programme were enrolled in Seguro Popular automatically by the state; therefore, for this group the campaign informed them of their new rights under Seguro Popular. The list of benefits in these areas is extensive and has been published,¹² so that intended beneficiaries would be aware of their rights in discussions with health clinics. The control health cluster in each pair received nothing extra and had the usual for-pay health care.

We published extensive details of our research design, baseline statistical analysis, and expectations of empirical results, including reports of large meetings with federal and state officials aimed to elicit and reveal their benchmarks for success,⁸ before analysing any outcome. Among other concerns, we anticipated heterogeneity in states' and localities' compliance with programme requirements and federal directives, and in individual-level non-compliance. Although state governments in our treatment areas undertook substantial media campaigns to encourage affiliation and to reduce barriers to affiliation by establishing programme offices, individuals were free to decide whether to affiliate.

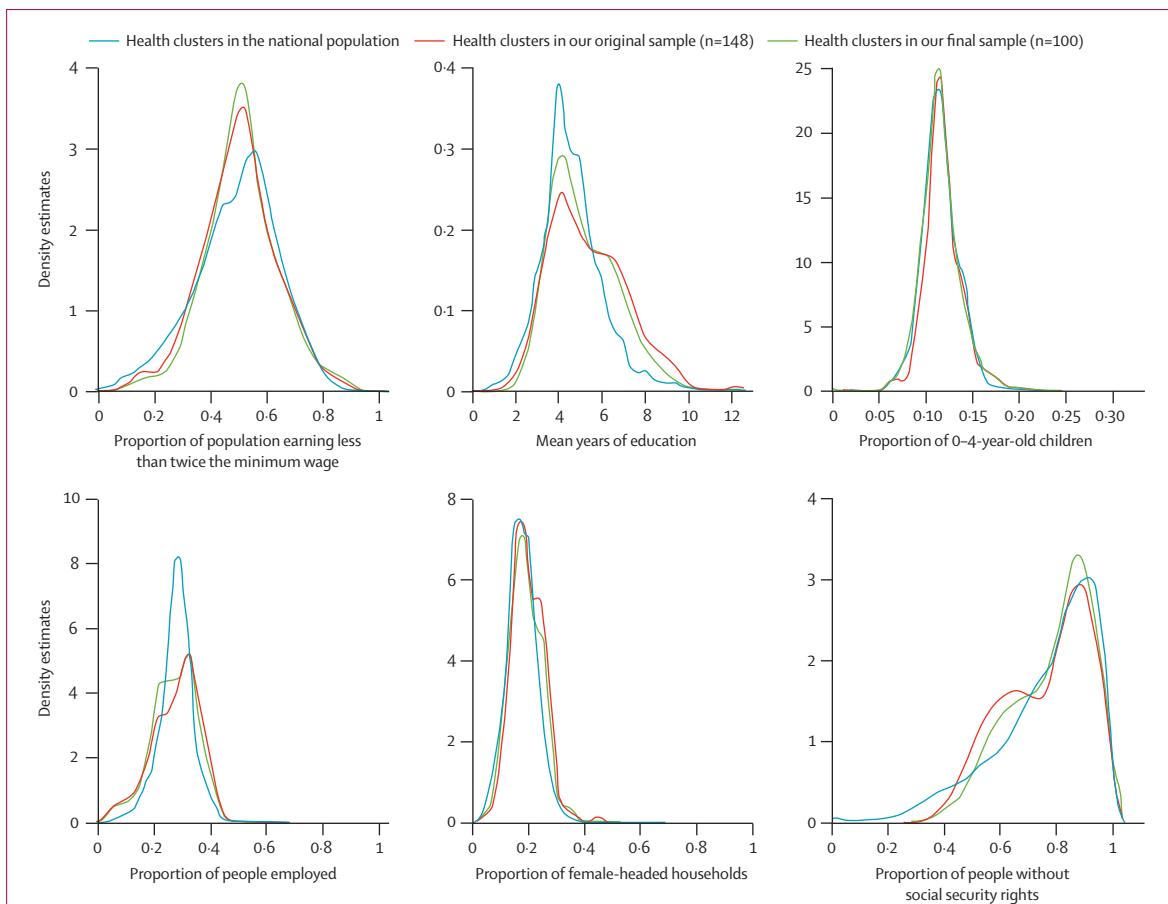


Figure 2: Comparisons between selected and all health clusters

Each histogram (density estimates) of health clusters indicates the national population (blue lines), our original sample of 148 (red lines), and our final sample of 100 (green lines). These histograms display the distribution of eligible cluster statistics, not the distribution for Mexican people or families.

Sample selection

Before the baseline survey, power calculations and administrative constraints indicated that we would be able to interview up to about 380 randomly chosen households in each of the 100 clusters. We selected 100 clusters from the original 148, emphasising within-pair similarity (to increase experimental control and statistical efficiency) and expected household compliance with the experimental protocol. Failure to pair before randomisation would have increased our standard errors by as much as 600%.¹³ Some clusters from each of the seven states are included: Guerrero (8), Jalisco (2), Estado de México (54), Morelos (26), Oaxaca (4), San Luis Potosí (4), and Sonora (2).

As with most clinical experiments, the set of clusters we could randomly assign to treatment was not a random selection from all clusters eligible for Seguro Popular nationwide, but was determined by constraints imposed by the timing of the experimental intervention and the ability of state governments to cooperate. We further narrowed the available clusters by a selection based on the closeness of the match and the likelihood of complying

with the experiment. Our population was similar to that eligible within the 12 284 original clusters identified for the experiment nationwide (10 616 rural and 1668 urban).

Figure 2 shows six demographic variables in the national population of health clusters eligible for the experiment ($n=12\,284$), our original sample of 148 health clusters, and our final sample of 100 health clusters. The differences for the proportion of the population earning less than twice the minimum wage, mean years of education, the proportion of children younger than 4 years of age, the proportion of people employed, the proportion of female-headed households, and the proportion of people without social security rights (ie, eligible to enrol in Seguro Popular), between the clusters were mostly small. Unfortunately, we did not have data for outcomes of interest for all clusters nationally, such as individual health and health spending outcomes, which would have provided clearer information on external validity.

Survey methods

We did a baseline survey at the time of random assignment that attempted to contact 36 181 randomly selected

See Online for webappendix

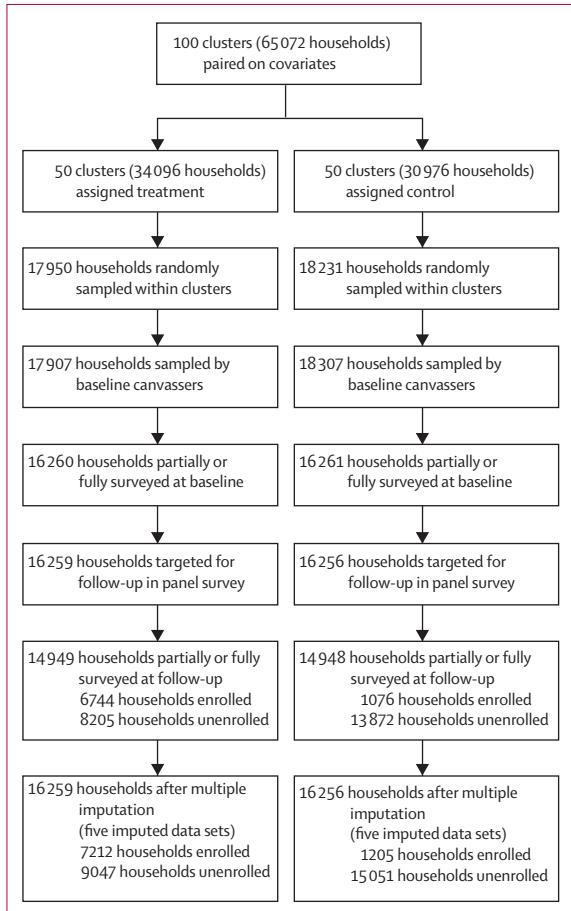


Figure 3: Flowchart describing randomisation protocol, sampling, survey panel attrition, and experimental compliance

Numbers are drawn from the household survey. Item-level non-response varied slightly for individual outcomes, and imputations done on multiple data sets produced slightly different imputations for programme affiliation totals.

households in the 100 selected clusters. We collected family spending data from the head of household and individual data from an adult randomly selected with a Kish table.¹⁴ A total of 32 515 households were at least partially surveyed in the baseline phase (a 90% response rate). We previously used this survey⁸ to validate the randomisation by verifying balance between treatment and control groups on background covariates and measures of the outcomes of interest at the time of random assignment, but before any treatment was administered. We found a pattern indicating potential bias in only one of our eight categories of variables: for the self-reported health variables, in the treatment group the poor tend to report that they are slightly healthier and the rich that they are sicker than those in the control group. Although the pattern is within what might be expected because of random chance, a placebo-like effect is also possible, and therefore we also adjusted for this by estimating the average treatment effect using difference-in-difference analysis (which had no effect for any other set of variables).

About 10 months later, we did a follow-up survey of households that we could reach during the baseline survey, fully or partially re-interviewing 29 897 households (figure 3). We defined a partially completed survey as one which obtained at least a household's Seguro Popular affiliation status at the follow-up survey.

Our major expenditure variables had only 1–2% missing values at baseline, and were only 8% missing in the assessment survey. We used standard multiple imputation techniques for missing values.^{15–17} These are described in the webappendix and were implemented as planned.⁸

Outcomes and quantities of interest

We identified outcomes of interest from the follow-up survey, focusing on expenditures—which was the main aim of Seguro Popular. Out-of-pocket health expenditures for all services, inpatient and outpatient care, medical devices, and medication were measured by annualising the head of household's self-reported out-of-pocket health spending in those categories during the previous 1–3 months. Catastrophic expenditures are defined on the basis of the share of a household's spending, after a minimal food budget that is allocated to health care. If a household's health spending exceeded 30% of the capacity to pay, the household was regarded as suffering catastrophic health expenditures. Other outcomes included self-reported health risk factors, health self-assessments, a blood pressure test and blood test for cholesterol, women's and maternal health variables (webappendix).

We examined these outcomes for several subgroups, as planned. We estimated causal effects for both low-asset and high-asset households. A household was defined as high-asset if it had at least half of the following items: a cement or tile (non-dirt) floor, electricity, washing machine, gas stove, refrigerator, phone, television, computer, and second home; and as low-asset otherwise. Other ways of aggregating these measures yielded similar conclusions. 55% of households in our sample fell into the high-asset category. The proportion of high-asset households in a pair is used as a cluster wealth measure. We also examined the programme's effect on female-headed households. We separately examined various individual-level outcomes using the same asset strata and, where appropriate, sex and age groups.

For every outcome variable, we present two causal quantities of interest. The first is an estimate of the intention-to-treat effect, which is the total effect of assigning a cluster to the programme (ie, upgrading medical facilities, providing insurance, encouraging households to enrol) regardless of experimental protocol compliance. The second quantity of interest is the complier average causal effect (CACE), which is the programme's effect on compliers, the (latent) group of individuals who would adhere to whatever treatment status they were randomly assigned to. CACE requires

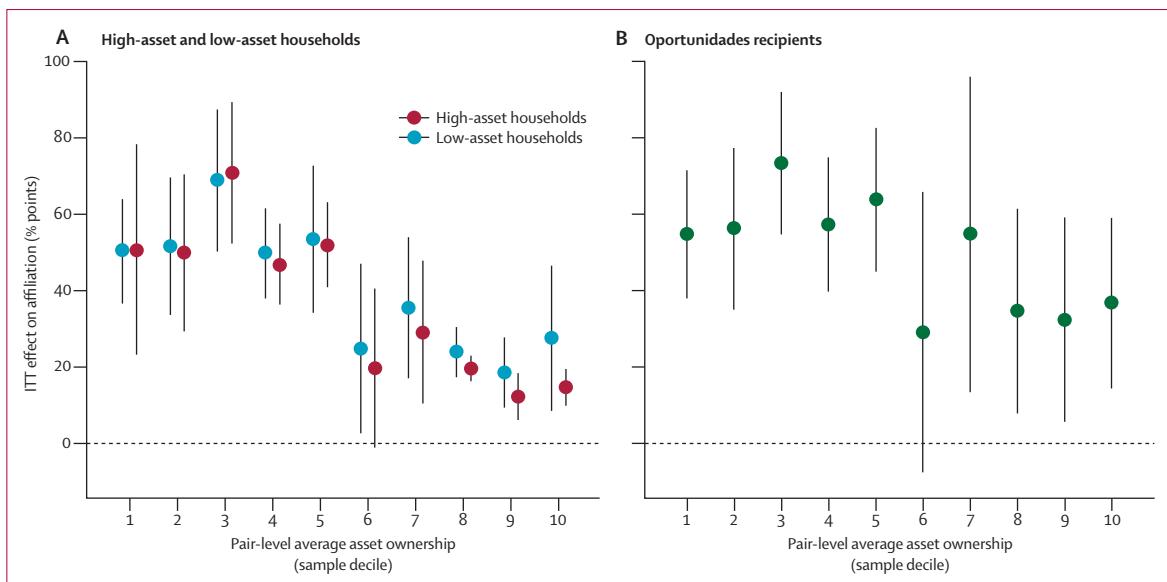


Figure 4: Effect of cluster-level assignment on affiliation to Seguro Popular for all households (A) and for those enrolled in the Oportunidades anti-poverty programme (B)

The horizontal axis in (A) and (B) indicates asset per person ownership deciles within our sample of matched pairs. The vertical axis in (A) and (B) represents the percentage point difference in self-reported Seguro Popular affiliation resulting from encouragement to affiliate. Points represent point estimates and the vertical bars are 95% CIs. ITT=intention to treat.

estimating and excluding from the analysis two types of households that by assumption do not benefit from the programme: never-takers—who are wealthy households and others who would not have enrolled in the programme regardless of their treatment assignment—and always-takers—typically highly motivated families who take advantage of the programme regardless of their treatment assignment. Estimation of intention to treat requires almost no statistical assumptions; therefore, we could use specially designed non-parametric methods.¹³ Estimation of CACE requires the assumption that only compliers benefit from Seguro Popular, which could be wrong in the long term but seemed reasonable for our 10-month assessment, and a few technical assumptions well-documented in the statistics literature¹⁸ and in the webappendix.¹³ All information needed to replicate the results in this report are publicly available.¹⁹

Role of the funding source

This project was partly supported by the Mexican Ministry of Health, which is responsible for Seguro Popular, and several authors were employed there. All data remained in the custody of Harvard University's Institute for Quantitative Social Science (IQSS), where researchers retained responsibility for data assembly and analysis.

The corresponding author was not employed by the ministry of health, had full access to the data, retained the right to publish results without previous review, and had final responsibility to submit for publication. Surveys were done by the National Institute of Public Health of Mexico, with funding from the ministry of health. IQSS provided other funding.

Results

A small number of households in control clusters enrolled in the programme, whereas many individuals in treatment clusters chose not to enrol. In control clusters, the self-report of affiliation rate was low (7%; 1205 of 16 256 households). In the treatment clusters, 44% (7212 of 16 259) of households reported being affiliated. We estimated that low-asset health clusters were composed of: 54% compliers, 37% never-takers (including 10–15% with pre-existing coverage), and 9% always-takers. In high-asset health clusters, these figures were 20%, 74%, and 6%, respectively.

We also estimated the effect of assignment on affiliation, stratifying by cluster socioeconomic level and household wealth level (figure 4). Results indicated that treatment assignment was much more effective in poorer areas than in areas with higher-asset ownership. Surprisingly, however, treatment assignment was not more effective for low-asset than for high-asset households. Our data suggest that uptake of the programme was mainly related to the wealth of the geographic area rather than of the individual family. Those with more assets living in poor areas are as likely to affiliate with Seguro Popular as their poorer neighbours.

Finally, we investigated households in our treatment clusters already enrolled in the Oportunidades anti-poverty programme, who had automatic access to Seguro Popular. We found that only 66% (5043 of 7628) of Oportunidades respondents in treated areas were aware of their automatic affiliation. Oportunidades recipients in poorer clusters were more likely to be aware of their affiliation than recipients in wealthier clusters (figure 4).

	All participants				Experimental compliers only			
	Control group*	ITT	SE	95% CI	Control group*	CACE	SE	95% CI
All	8·4	1·9†	0·9	0·2 to 3·7	9·5	5·2†	2·3	0·8 to 9·6
Low asset	9·9	3·0†	1·3	0·5 to 5·5	11·0	6·5†	2·5	1·6 to 11·3
High asset	7·1	0·9	0·8	-0·7 to 2·5	7·9	3·0	2·7	-2·3 to 8·4
Female-headed	8·5	1·4	1·1	0·7 to 3·5	9·6	3·8	3·0	-2·1 to 9·7

ITT=intention to treat. SE=standard error. CACE=complier average causal effect. *Average value of the variable, as a baseline. † $p\leq 0·05$ (one-tailed t test). Positive values correspond to favourable outcomes. Catastrophic expenditures are defined as out-of-pocket health expenses greater than 30% of post-subsistence income.

Table 1: Causal effect of rolling out Seguro Popular on the reduction in the number of households suffering catastrophic health expenditures

During the 10-month assessment period, Seguro Popular reduced by 23% the proportion of all respondents experiencing catastrophic expenditures (table 1). Among compliers, Seguro Popular reduced the proportion of catastrophic expenditures by 55%. Most of the overall effect occurred in poorer households. Effects for high-asset families and female-headed households were smaller but still positive, and measured with about the same precision.

Table 2 shows the causal effect of Seguro Popular on out-of-pocket expenditures for health care. Average out-of-pocket health expenditures in the control group were about MXN\$1631 overall, with higher-asset households spending more on care than low-asset households. Among experimental compliers, the effects were about double the intention-to-treat effect. High-asset and female-headed households had effects that were slightly smaller than those for low-asset households, albeit less precisely measured. Seguro Popular also reduced out-of-pocket expenses for inpatient and outpatient medical care, overall and for poor households, but had no detectable positive effect on spending for medication and medical devices.

We assessed the effects of Seguro Popular on various other outcomes: use of medical facilities, modifiable risk factors, measures of the diagnosis and coverage for chronic illness, and health self-assessments (table 3).

Before the study was started, government officials had expressed concern about the short assessment period, but expected that the programme would produce at least some effect on the use of regular and preventive medical services at this initial stage.⁸ Our findings do not confirm these expectations or those in related previous observational research.^{6,7,20} We found that Seguro Popular had no substantial effect on use of medical services, regardless of how it was measured and the subgroup studied. Indeed, most point estimates are negative, indicating a slight drop in services in the treated compared with control clusters. Separate subgroup analyses for low-asset, high-asset, and female-headed households also showed no significant effects (webappendix).

We also estimated satisfaction with the programme. Among Seguro Popular enrollees, 69% rated the quality of health services as very good or good (as opposed to moderate, bad, or very bad), and 85% reported that programme benefits were explained by programme officials in a manner that was clear or very clear. Furthermore, 94% reported that they were treated well during the affiliation process and 97% planned to enrol again in the follow-up period.

Finally, we examined the effects of Seguro Popular on nine separate health self-assessments, by both intention to treat and CACE. Seguro Popular effects on these outcome variables were positive and measured with precision. Unfortunately, but as anticipated, this effect was only apparent because slightly positive effects

	All participants			Experimental compliers only		
	Control group*	ITT	SE	Control group*	CACE	SE
Overall						
All	1631·3	258·0	175	1712·7	689·7	453
Low asset	1360·2	425·6†	197	1502·6	915·3†	392
High asset	1867·9	128·4	201	1933·2	428·2	669
Female-headed	1509·1	156·5	207	1535·0	428·6	566
Inpatient care						
All	532·5	96·9†	44	557·1	259·1†	112
Low asset	527·1	188·2†	73	579·0	404·8†	142
High asset	537·2	31·1	52	536·2	103·6	173
Female-headed	452·5	115·1†	68	463·3	315·2†	182
Outpatient care						
All	448·3	116·7†	63	499·1	312·0†	161
Low asset	412·3	176·7†	73	466·3	380·0†	147
High asset	479·7	81·9	69	533·0	272·9	230
Female-headed	416·3	110·4	75	451·3	302·4	202
Medicine						
All	521·1	20·0	41	534·5	53·3	109
Low asset	427·3	17·8	46	444·7	38·3	100
High asset	603·0	29·4	47	627·5	98·1	157
Female-headed	625·6	53·6	55	671·2	146·8	151
Medical devices						
All	139·7	-8·8	23	117·8	-23·4	62
Low asset	72·0	-0·2	20	72·8	-0·5	43
High asset	198·8	-16·5	29	165·6	-55·1	98
Female-headed	155·5	10·9	34	147·9	30·0	94

ITT=intention to treat. SE=standard error. CACE=complier average causal effect. *Average value of the variable, as a baseline. † $p\leq 0·05$ (one-tailed t test). Values are expressed in MXN\$. Positive values correspond to favourable outcomes. Inpatient care includes self-reported spending on staying overnight in a hospital or health clinic, whereas outpatient care includes spending on medical care that did not require an overnight stay. Medicine includes spending on traditional and non-traditional medications, whereas medical devices include glasses, prostheses, hearing aids, and others.

Table 2: Causal effect of Seguro Popular on the self-reported out-of-pocket health expenditures

The primary purpose of our assessment was to estimate the effect of Seguro Popular on the percentage of households with catastrophic health expenditures, as well as out-of-pocket costs. Table 1 shows that 8·4% of our control group sample spent more than 30% of their post-subsistence income (ie, capacity to spend after a minimal food budget) on health care in the past year.

appeared even in the baseline survey.⁸ To check whether the results at 10 months were affected by these placebo differences, we estimated the average treatment effect of individuals' difference between the follow-up period and baseline. With this difference-in-difference analysis, the average effects of the programme on health self-assessments on individuals in various subgroups, including high-asset and low-asset households and in women, are small and close to zero. This result emphasises the importance of a baseline survey in protecting from placebo effects even in large-scale non-double-blind public policy assessments.

Discussion

We assessed a new public policy that increases health-care funding to Mexican states. It grants them additional administrative responsibility while improving federal supervision, and it provides individuals with health insurance and information resources. Although the new stewardship model of health-care delivery that Seguro Popular intended to implement^{4,21} (compared with previous centralised and decentralised systems) has been supported by conceptual and observational studies, no country pursuing such a model in its health sector has previously done a large-scale randomised scientific assessment.

That Mexico may now have a validated architecture for delivering health services to the poor seems to be the most surprising and encouraging result from this experiment. We found that the stewardship of Seguro Popular has been successful in reducing overall catastrophic and out-of-pocket expenditures for inpatient and outpatient medical procedures, especially in the poorest individuals.

We identified several issues with Seguro Popular that could indicate implementation challenges or other difficulties in the design of the policy. In our sample, after 10 months the programme had neither reduced spending on medication nor increased health-care use or diagnoses. Our surveys do not ascertain whether the result means that there was no effect at all, whether the reduction in price did not lead to an increase in quantity of drugs, or whether a change will arise over a period longer than 10 months. Although neither we nor large groups of federal and state experts, political appointees, and programme officials anticipated this finding, the result might be explained by the short assessment period, similar to the experience in other areas that adopt new health programmes (eg, in Massachusetts, a recently implemented universal health insurance programme is also associated with difficulties in finding a family primary care physician and an increase in the number of physicians refusing new patients).²² The affiliation programme could be improved, especially in the enrolment of poor families living in relatively wealthy communities. Similarly, treatment-cluster households in the Oportunidades anti-poverty pro-

	All participants			Experimental compliers only		
	Control group*	ITT	SE	Control group*	CACE	SE
Utilisation (procedures)						
Used outpatient services	62.6%	-1.5%	1.9%	64.8%	-4.0%	5.2%
Outpatient visits (count)	1.6	-0.03	0.09	1.7	-0.08	0.23
Hospitalised	7.6%	-0.2%	0.5%	7.9%	-0.5%	1.3%
Hospitalisations (count)	0.1	-0.003	0.006	0.1	-0.01	0.02
Satisfaction with provider	68.0%	-1.0%	1.6%	69.8%	-2.6%	4.5%
Utilisation (preventative)						
Eye exam (past year)	10.0%	-0.7%	0.7%	9.8%	-1.8%	1.9%
Flu vaccine	25.7%	-1.8%	1.4%	27.2%	-4.9%	3.7%
Mammogram (past year)	5.1%	-0.9%	0.6%	5.2%	-2.3%	1.6%
Cervical (past year)	21.8%	-1.3%	2.0%	22.2%	-3.2	4.8%
Pap test (past year)	31.9%	-2.3%	2.1%	33.2%	-5.8%	5.0%

*Average value of the variable, as a baseline. Data are percentages, unless otherwise indicated. ITT=intention to treat. SE=standard error. CACE=complier average causal effect.

Table 3: Causal effect of Seguro Popular on utilisation of medical procedures and preventive care

gramme were affiliated automatically, but a third of them were not aware of this fact. Our assessment has thus given the federal stewards of Seguro Popular some of the knowledge they need to further improve and assess the programme.

After randomisation, the first step to complete the intervention was to affiliate the eligible population to Seguro Popular. We did this successfully, reaching an affiliation rate of 44% in intervention areas. These high rates of affiliation made an immediate impact on out-of-pocket expenses possible, because beneficiaries used health-care facilities at no cost. The no-charge policy became effective immediately after affiliation, which made it possible for us to document a significant reduction in the prevalence of households with catastrophic health expenditure. However, other components of the operation of Seguro Popular require complex administrative actions, before rollout is complete. For example, the improvement of drug supplies required purchase using open bids that, in the best of cases, took 6–8 months to be completed. In the same way, the process of hiring and training new medical staff willing to work in these remote clinics and the eventual accreditation of medical facilities required more time than was available during our assessment period.

Results from this study differ from those of observational studies for several reasons. The absence of an effect on medication expenditures is surprising because of the aims of the programme and past observational results, which found that Seguro Popular families spend about 14% less on drugs than uninsured families,⁶ and suggested that the programme had made drugs more available and affordable.⁹ Other observational studies had reported that the number of hospital discharges increased with overall Seguro Popular

enrolment, and effective coverage rates for various medical services were higher for Seguro Popular individuals than for those uninsured.⁶ Some of these differences might arise from the short assessment period, different observation periods or target populations, or from bias in the observational studies that our ability to randomise enables us to avoid.

Continued assessment of the programme is needed. Furthermore, we could only make inferences about our study areas and similar communities. Expanding the study to include parallel experiments with new groups in additional health clusters in other parts of the country would strengthen generalisations to the nation. Finally, although Seguro Popular is unprecedented in scope, we could only test the effects of the programme as a whole rather than each component. We hope that other researchers will continue this research and apply our experimental design to other public policy reforms, both in Mexico and elsewhere.

Contributors

All authors made contributions to each part of the project, participated in extensive planning and design meetings, and helped to interpret the findings. GK had overall responsibility for the project and drafted the manuscript. GK, EG, MHA, and HHL participated in all phases of the design and implementation. EG was responsible for the survey questionnaires. GK, KI, RTM, and CN developed the statistical methods. GK, KI, RTM, CN, NR, and EG analysed the data. JL did qualitative research and studied the history and implementation of the reform. MV coordinated the research teams and did data quality checks. MMTR, JEHA, and MHA constructed the health clusters and did survey planning and implementation. All authors contributed to revisions of the manuscript and approved the final version.

Conflicts of interest

This study was partly supported by the Mexican Ministry of Health under the previous administration. GK, EG, KI, JL, RTM, CN, and NR have no conflict of interest and retained the right to publish without previous review or restriction. MV and MHA are employed by the Mexican Ministry of Health. MMTR and JEHA work at the National Institute of Public Health. HHL has been National coordinator of the Seguro Popular programme.

References

- 1 Gwatkin DR, Wagstaff A, Yazbeck AS, eds. *Reaching the Poor*. Washington, DC: The World Bank, 2005.
- 2 Frenk J. Sistema de Protección Social en Salud, Elementos conceptuales, financieros, y operativos. Mexico City: Secretaría de Salud, 2005.
- 3 Frenk J, Sepúlveda J, Gómez-Dantés O, Knaul F. Evidence-based health policy: three generations of reform in Mexico. *Lancet* 2003; **362**: 1667–71.
- 4 Knaul FM, Frenk J. Health Insurance in Mexico: achieving universal coverage through structural reform. *Health Affairs* 2005; **24**: 1828–41.
- 5 González-Pier E, Gutiérrez-Delgado C, Stevens G, et al. Priority setting for health interventions in Mexico's System of Social Protection in Health. *Lancet* 2006; **368**: 1608–18.
- 6 Gakidou E, Lozano R, González-Pier E, et al. Assessing the effect of the 2001–06 Mexican health reform: an interim report card. *Lancet* 2006; **368**: 1920–35.
- 7 Frenk J. Bridging the divide: global lessons from evidence-based health policy in Mexico. *Lancet* 2006; **368**: 954–61.
- 8 King G, Gakidou E, Ravishankar N, et al. A 'politically robust' experimental design for public policy evaluation, with application to the Mexican universal health insurance program. *J Policy Analysis Manag* 2007; **26**: 479–506.
- 9 Hussey MA, Hughes JP. Design and analysis of stepped wedge randomized trials. *Contemp Clin Trials* 2007; **28**: 182–91.
- 10 Imai K, King G, Stuart E. Misunderstandings among experimentalists and observationalists about causal inference. *J R Stat Soc, Series A* 2008; **171**: 481–502.
- 11 Moore RT. Political analysis and statistical applications for social policy research. Harvard University, 2008.
- 12 Mexican Ministry of Health. Seguro Popular Benefits. http://www.seguro-popular.salud.gob.mx/contenidos/menu_beneficios/beneficios_inicio.html (accessed Feb 28, 2009).
- 13 Imai K, King G, Nall C. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Stat Science* 2009. In press.
- 14 Kish L. *Survey sampling*. New York: John Wiley, 1965.
- 15 Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley, 1987.
- 16 King G, Honaker J, Joseph A, Scheve K. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *Am Political Science Rev* 2001; **95**: 49–69.
- 17 Honaker J, King G. What to do about missing values in time series cross-section data. <http://gking.harvard.edu/files/abs/pr-abs.shtml> (accessed Sept 6, 2008).
- 18 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–55.
- 19 King G, Gakidou E, Imai K, et al. Replication data for: public policy for the poor? A randomised ten-month evaluation of the Mexican universal health insurance program. <http://hdl.handle.net/1902.1/11044>.
- 20 Nigenda G, Orozco E, Olaiz G. La importancia de los medicamentos en la operación del Seguro Popular de Salud. In: Knaul F, Nigenda G, editors. *Caleidoscopía de la Salud*. Funsalud, 2003; 263–73.
- 21 Centro de Estudios de las Finanzas Públicas, Cámara de Diputados, "Gasto en el Sector Salud", Nota Informativa, 64, Sept, 2006. <http://www.cefp.gob.mx/notas/2006/notacefp0642006.pdf> (accessed Feb 21, 2009).
- 22 Massachusetts Medical Society, Physician Workforce Study: full report 2008. <http://www.massmed.org/AM/Template.cfm?Section=Home&TEMPLATE=/CM/ContentDisplay.cfm&CONTENTID=23166> (accessed Feb 21, 2009).