Public Policy for the Poor? A Randomised Assessment of the Mexican Universal Health Insurance Program

Ryan T. Moore
Washington University in St Louis

Joint work with Gary King, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas

World Bank/SIEF Impact Evaluation Workshop for Health Sector Reform
Cape Town, South Africa
9 December 2009

Seguro Popular: A Massive Reform

- Intervention: resources for medical services, preventive care, pharmaceuticals, access, and financial health protection
- Beneficiaries: 50M Mexicans (half the population) without regular access to health care, particularly low incomes, informal sector work.
- Premiums: $0 for bottom 2 income deciles, low otherwise
- Cost in 2005: $795.5 million in new money
- Cost when fully implemented: additional 1% of GDP $(5.6\% \rightarrow 6.6\%)$
- One of the largest health reforms of any country in last 2 decades
- Most visible accomplishment of the Fox administration
- Major issue in the 2006 presidential campaign (and evidence of impact on 2006 vote choice: 7-11% more likely for Calderón)

Goals of SP & Evaluation Outcome Measures

- Financial Protection (money for the poor rarely makes it there)
  - Out-of-pocket expenditure
  - Catastrophic expenditure (8.4% of households, & 10% of the poor, spend $> 30\%$ of annual disposable income on health)
  - Impoverishment due to health care payments
- Utilization
  - Preventative
  - Inpatient, outpatient
- Health
  - Health status
  - All-cause mortality
  - Cause-specific mortality
- Health System Effective Coverage
  - Percent of population receiving appropriate treatment by disease
  - Responsiveness of Seguro Popular
  - Satisfaction of affiliates with Seguro Popular
- Health Care Facilities
  - Operations, office visits, emergencies, personnel, infrastructure and equipment, drug inventory.

SP Evaluation

- Health Minister and President asked: How can one democratically elected government "tie the hands" of its successors?
  - Commission an independent evaluation
  - As in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: no gain in extending it
- The largest randomized health policy experiment in history
- One of the largest policy experiments to date
- First cohort: 148 geographic areas, 1,380 localities, $\approx 118,569$ households, $\approx 534,457$ people

Lessons from Previous Public Policy Experiments

- Most large scale public policy experiments fail – unnecessarily!
- Many failures are political
  - politicians: need to pursue short term goals
  - local political actors: want local results
  - citizens: you plan to randomly assign me?
  - budgets & priorities: can change midstream
  - all perfectly legitimate; natural consequences of democracy
- E.g., Oportunidades program: Some governors found money for control groups to participate
  (numerous similar examples worldwide – US, Kenya, UK, Sweden)
- Previous evaluation designs ignored democratic politics
- We developed a new research design & new methods:
  - includes fail-safe components for when politics intervenes
  - uses data far more efficiently to find effects and save money

Example of Fail-Safe Design Procedure (CR vs. MPR)

1. Complete Randomization (used in Oportunidades evaluation)
   - Flip coin to assign program to each area
   - If one area is lost:
     - treated and control groups are incomparable
     - advantages of randomization are gone
2. Matched-Pair Randomization (used in Seguro Popular evaluation)
   - Match areas in pairs on background characteristics
   - Flip coin once for each pair: one area within each pair gets the program
   - If one area is lost:
     - Drop the other member of the pair
     - Remaining pairs are kept
     - Treated and control groups are still protected by randomization: advantages of the experiment survives
   - With our new statistical methods, the design:
     - More efficient: up to 38 times!
     - Smaller standard errors: up to 6 times smaller
     - We can find effects where complete randomization cannot
     - Far less expensive for the same impact

Detailed Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics.
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)
5. Randomly assign one in each pair to receive encouragement to affiliate, better health facilities, drugs, and doctors
6. Conduct baseline survey of each cluster's health facility
7. Survey $\approx$32,000 random households in 50 of the 74 treated and control unit pairs (chosen based on likelihood of compliance with encouragement and similarity of the clusters within pair)
8. Implement SP
9. Repeat surveys in 10 months to measure effects

Matched-Pair Cluster-Randomized Designs, I

- MPDs $\geq$ Complete Randomization w.r.t.:
  - efficiency
  - bias
  - power
  - estimator simplicity
  - robustness to political intervention
- Vast majority of political science CREs ignore MPD advantages (hopefully changing: Kosovo election security [fall 2009], China health insurance packages [26 Nov 09!], . . .)
- Conclusion: avoid leaving useful information on the table!

Matched-Pair Cluster-Randomized Designs, II

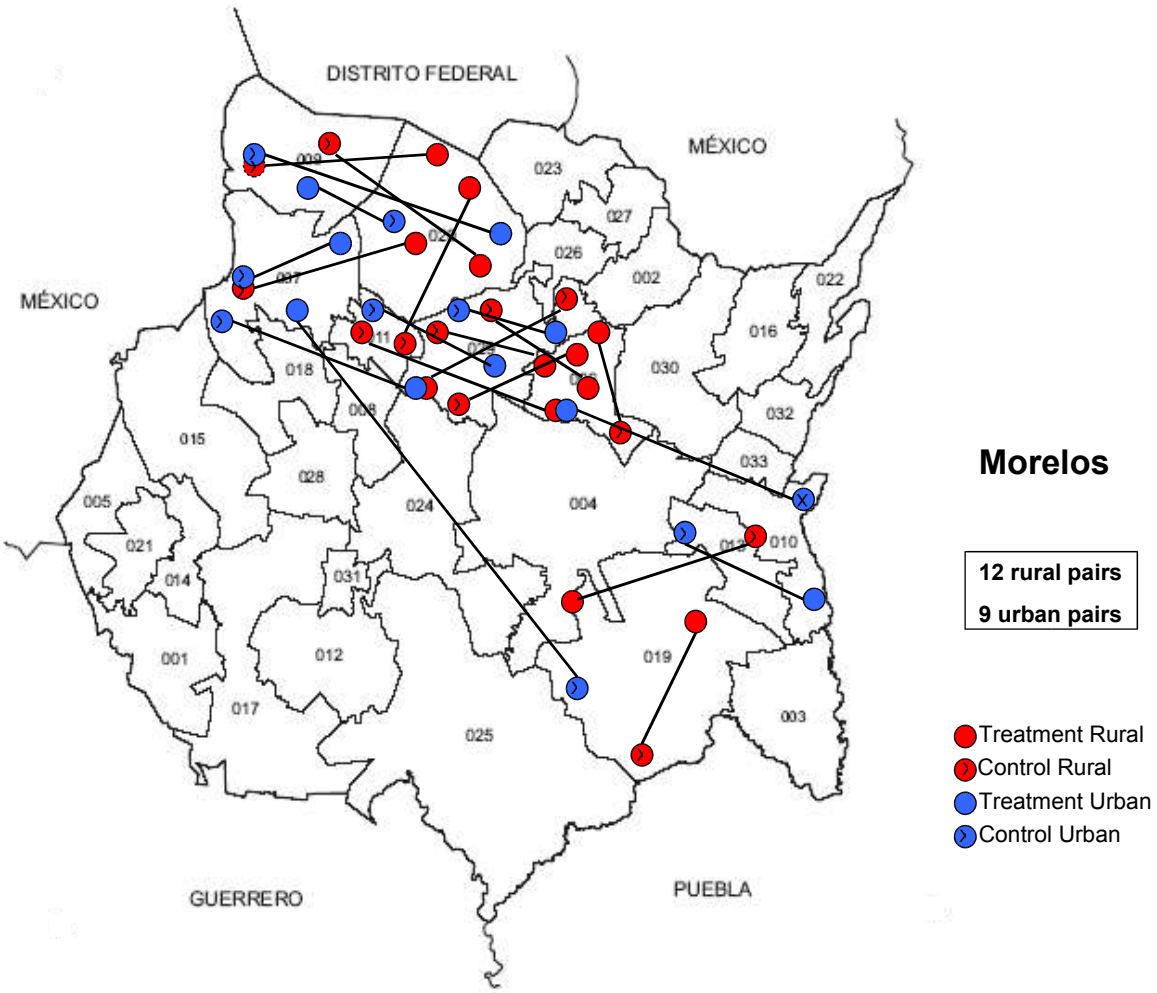
- Data on clusters:

cluster	population	educ years	doctors	nurses
MCSSA000364	3130	4.00	1	1
MCSSA000504	6492	5.11	1	1
MCSSA008221	5096	4.26	2	2

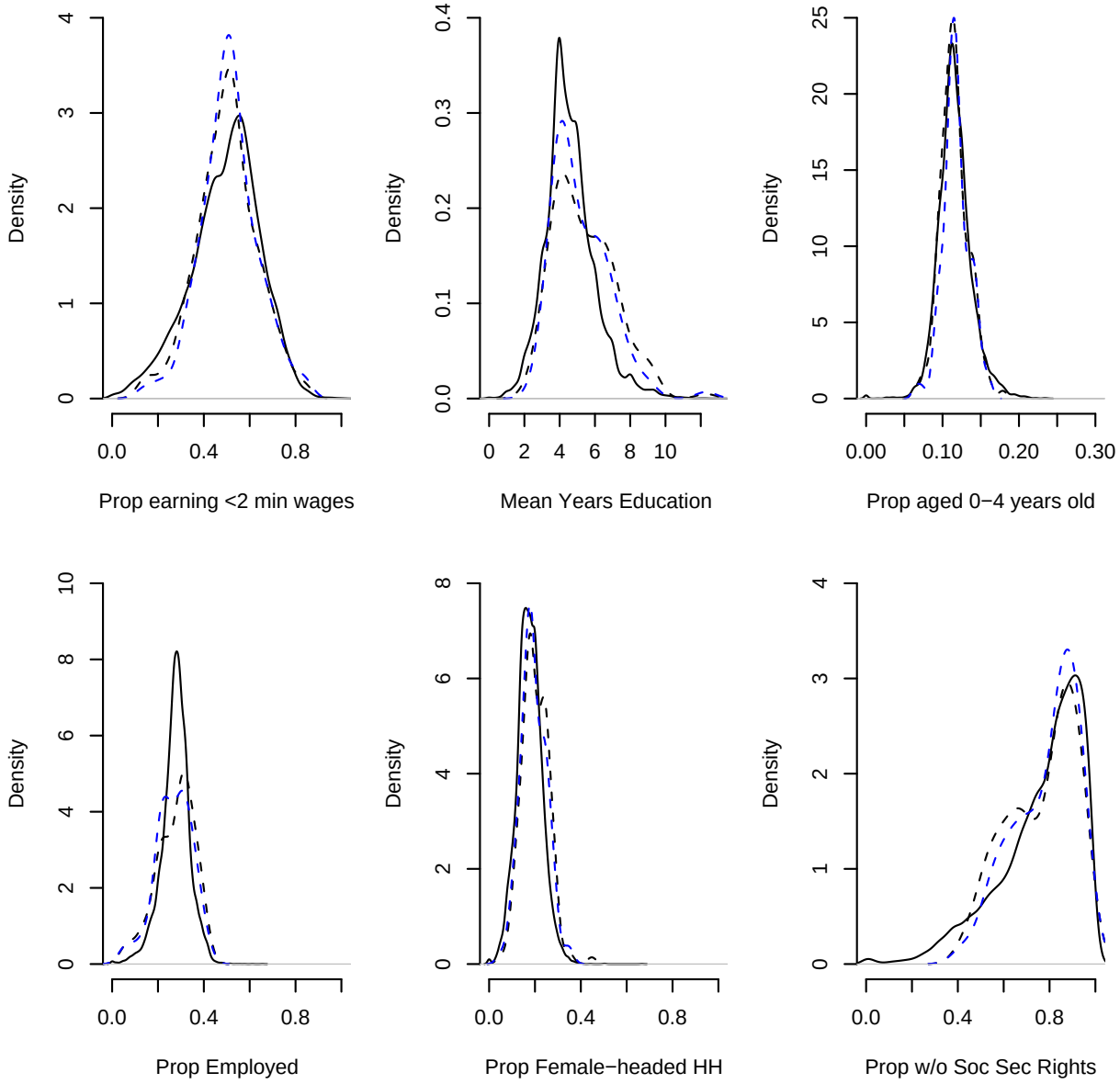
- Calculate Mahalanobis distance between every pair of units: $MD_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)'\Sigma^{-1}(\mathbf{x}_i - \mathbf{x}_j)}$ (or resistant alternative)
- Use paired distances to created matched sets
- Other constraints: within states, (no more than 1 year of educ apart, ...)
- `blockTools` in `R` (Moore 2009)

SP Evaluation: 148 clusters (74 pairs) in 7 states

Matched Pairs, Morelos



Clusters are Representative On Measured Variables



Design and Analysis Strategy is Triply Robust

Design has three parts

1. Matching pairs on observed covariates
2. Randomization of treatment within pairs
3. If necessary, statistically adjust for differences

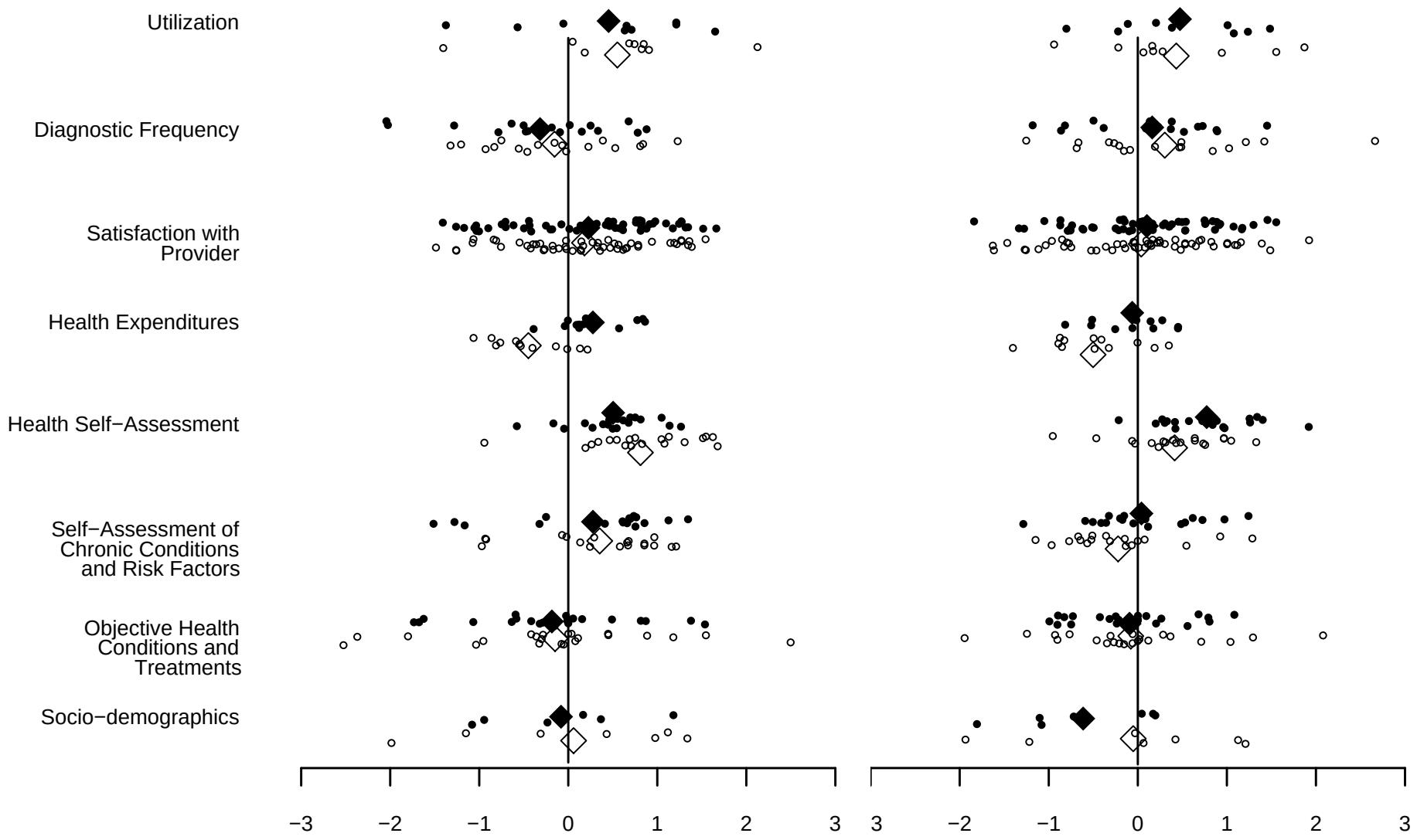
Triple Robustness

If matching or randomization or statistical analysis is right, but the other two are wrong, results are still unbiased

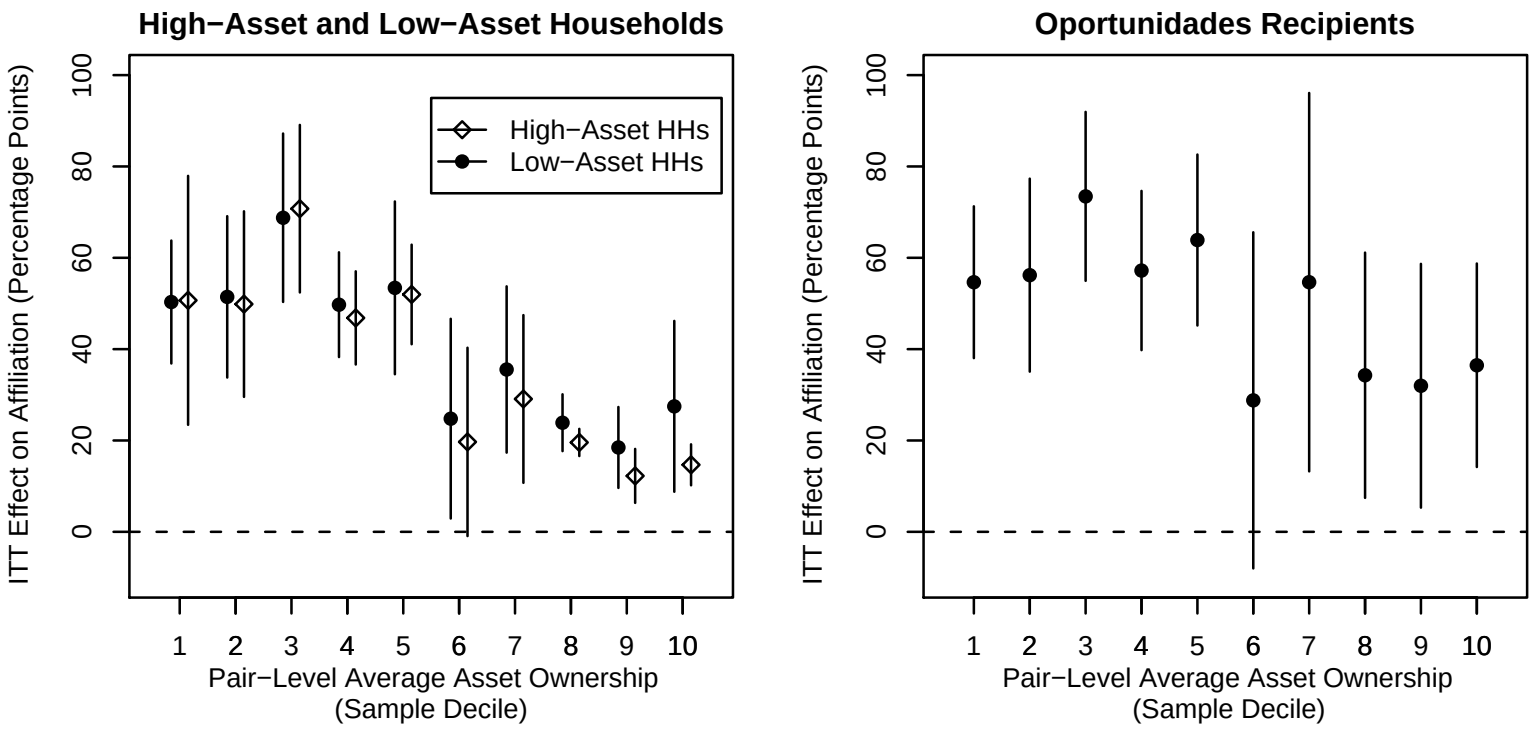
Two Additional Checks if Triple Robustness Fails

1. If one of the three works, then "effect of SP" on time 0 outcomes (measured in baseline survey) must be zero
2. If we lose pairs, we check for selection bias by rerunning this check

ITT on Outcome Measures at Baseline, for all families (left) and poor families, in Oportunidades (right)



Effect of Encouragement on Seguro Popular Affiliation



Horizontal axes: per-capita asset ownership deciles of areas (poorer to left).
Vertical axes: percentage point causal effect of encouragement to affiliate on SP affiliation. Affiliation $\approx 50 - 70\%$ for HH in below-median asset clusters.
Poor areas, not poor households, are affiliated the most.

Effect on % of Households with Catastrophic Health Expenditures

	All Study Participants			Experimental Compliers		
	Average (Control)	ITT	SE	Average (Control)	CACE	SE
All	8.4	1.9*	(.9)	9.5	5.2*	(2.3)
Low Asset	9.9	3.0*	(1.3)	11.0	6.5*	(2.5)
High Asset	7.1	0.9	(0.8)	7.9	3.0	(2.7)
Female-Headed	8.5	1.4	(1.1)	10.6	3.8	(3.0)

"Catastrophic expenditures": out-of-pocket health expenses $> 30\%$ of post-subsistence income. Percentage reductions:

	All	Compliers
All	$1.9/8.4 = 23\%$	$5.2/9.5 = 55\%$
Low Asset	$3/9.9 = 30\%$	$6.5/11 = 59\%$

Effect on Out-of-pocket Health Expenditures, I (in pesos)

	All Study Participants			Experimental Compliers		
	Average (Control)	ITT	SE	Average (Control)	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)	1689.9	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)	510.0	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)	496.8	302.4	(202)

Effect on Out-of-pocket Health Expenditures, II (in pesos)

	All Study Participants			Experimental Compliers		
	Average (Control)	ITT	SE	Average (Control)	CACE	SE
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)	738.9	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)	162.8	30.0	(94)

Conclusions

- Positive effects detected now:
  - Catastrophic expenditures slashed
  - In-patient out-of-pocket expenditures drastically reduced
  - Out-patient out-of-pocket expenditures drastically reduced
  - Citizen satisfaction is high
- Positive effects not yet seen:
  - Expenditures on medicines
  - Utilization (preventative, inpatient, outpatient, procedures)
  - Health risk factors
- Other findings:
  - Only 66% of automatically affiliated Oportunidades respondents were aware of this fact
  - More encouragement to affiliate might be devoted to finding the poor hidden within relatively "wealthier" clusters
  - Developed new, more powerful evaluation design and statistical methods
  - Seguro Popular evaluation design: being copied around the world

Citations

King, Gary, 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 and Héctor Hernández Llamas. "Public Policy for the Poor? A Randomized Assessment of the Mexican Universal Health Insurance Program." *The Lancet*, 373(9673):1447–54, 25 April 2009.

Replication data, Dataverse: `http://dvn.iq.harvard.edu/dvn/dv/rmoore`

Moore, Ryan T. `blockTools: Blocking, Assignment, and Diagnosis for Randomized Experiments`, R package version 0.4-1. `http://rtm.wustl.edu/software.blockTools.htm`. 2008-09.

Kosuke Imai, Gary King, and Clayton Nall. "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health Insurance Evaluation." *Statistical Science*, 24(1): 29–53, 2009.

Moore, Ryan T. **Political Analysis and Statistical Applications for Social Policy Research**. Ph.D. Thesis, Harvard University, May 2008.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila and Héctor Hernández Llamas. "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management*, 26(3): 479–509, 2007.