

Evaluation of Seguro Popular: Baseline Analysis

Gary King
Institute for Quantitative Social Science
Harvard University

Thanks, in Mexico, to Edith Arano, Carlos Avila, Juan Eugenio Hernández Avila, Mauricio Hernández Avila, Jorge Carreon Manuel Castro, Octavio Gómez Dantes, Sara Odette León, Héctor Hernández Llamas, René Santos Luna, Tania Martínez, Gustavo Olaíz, Maritza Ordaz, Eduardo González Pier, Héctor Peña, Raymundo Pérez, Esteban Puentes, Corina Santangelo, Sergio Sesna, Sara Uriega, Martha María (Mara) Tellez; at Harvard, to Dennis Feehan, Emmanuela Gakidou, Jason Lakin, Diana Lee, Ryan T. Moore, Nirmala Ravishankar, Manett Vargas, and our Panel of Experts, Edmundo Berumen, Luis Felipe Lopez Calva, Nora Claudia Lustig, Thomas Mroz, John Roberto Scott.

The First Results of our Evaluation (Effect of Random Assignment on One Mexican)

Before Treatment



After Treatment



(Manett's) Arturo Vargas

Evaluation Components

- **Impact Evaluation** (today's talk)
- National Level Analysis
- Process Evaluation
- In-depth Focus Groups

Goals of SP & Evaluation Outcome Measures

- Financial Protection
 - Out-of-pocket expenditure
 - Catastrophic expenditure (now 3% of households spend > 30% of disposable income on health)
 - Impoverishment due to health care payments
- 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.
- Health
 - Health status
 - All-cause mortality
 - Cause-specific mortality

- Panel survey ($n = 36,000$) at time 0 and 10 months later
- Aggregate data describing health clinics and areas around them
- Health facilities survey
- Focus group interviews

Quantities of Interest, for Each Outcome Variable

- Effect of rolling out the policy in an area (“intention to treat”)
 - Affiliating the poor automatically
 - Establishing an MAO, so people can affiliate
 - Encouraging others to affiliate: painting buildings, radio, TV, loudspeakers, etc.
- Effect of one Mexican affiliating with SP (“treatment effect”)
- Compliance rates:
 - Difference between intention to treat and treatment
 - A measure of program success
- Variation in effect size
 - Areas with no health facilities: SP effect zero
 - People who already have access to health care: SP effect small
 - Places with better doctors and health administration: bigger effects

Ideal Design for Mexican Society

- Roll out SP as fast as possible to as many as possible
 - Unless SP doesn't work!
 - Unless we can improve outcomes by learning from sequential affiliation
- Immediately give all Mexicans equal ability to affiliate
 - **Impossible**: insufficient health facilities in some areas
 - **Politically Infeasible**: local officials want benefits for their favored areas first

How “Ideal Designs” Make Evaluation Hard

- If anyone can affiliate
 - The older and sicker will affiliate first
 - Younger and healthier will affiliate less
 - I.e., affiliates are sicker than non-affiliates
 - Evaluation: affiliating makes you sick!
 - This is the problem of “selection bias”
- If politicians (in a democracy) decide which areas get MAOs
 - Privileged areas get affiliation first
 - Political favorites are affiliated early
 - Even if SP has no effect, areas with SP will be healthier

A Feasible Design for Scientific Evaluation

First Define and Choose Health Clusters

- Divide country into “health clusters”
 - Clínicas, centros de salud, hospitales, etc., and catchment area
 - Catchment area based on time to service
 - Rural clusters: set of localidades that use the health unit.
 - Urban clusters: set of AGEB’s that use the health unit.
- Reasons to exclude areas from evaluation
 - **Political**: politicians want favorite areas covered; some don’t want their states participating in the evaluation
 - **Institutional**: Drop (rural) clusters without adequate facilities
 - **Administrative**: Drop (rural) clusters with < 1000 population; Only include urban clusters with 2,500–15,000 population
 - **Methodological**: Drop areas where affiliation had already started

Remaining in study: 148 clusters in 7 states



States and Clusters **not** Selected Randomly

- Effect of SP on the areas studied
 - estimated well (using methods to be described)
- Ways to Estimate Effects of SP on all of Mexico
 - Assume constant effects: probably wrong
 - Hints from present study: how effects of SP varies due to geography, income, age, sex, etc.
 - Extrapolation: entirely model dependent
 - **Our strategy: Repeat design in other areas**
 - Same strategy as in most medical intervention studies

Who Can Affiliate?

Constraints

- Must choose clusters to roll out program, and
 - Affiliate the poor automatically
 - Establish an MAO, so people can affiliate
 - Encourage people to affiliate: radio, TV, loudspeakers, knock on doors, paint buildings, etc.
- Financial constraints: rollout must be staged over time

Randomized Evaluation Design

- Randomly select half of the 148 clusters for encouragement
- Other clusters to get encouragement at a later date
- Any Mexican family may still affiliate at any time
- No randomization at individual level
- Without an evaluation, choices would still be made, but would be arbitrary choices made by local government officials

Classical Randomization is Insufficient in the Real World

- Goal: equivalent treatment and control groups
- Classical random assignment achieves equivalence:
 - on average (or with a large enough n), and
 - if nothing goes wrong
- But, if we lose clusters
 - Equivalence of affiliate and non-affiliate clusters could fail
 - E.g., maybe poor, unhealthy clusters are more likely to drop out
 - Consequence: Bias in evaluation conclusions
- We need estimators robust not merely to statistical assumptions but to real world problems

We Use: Paired Matching, then Randomization

Design

- Sort 148 health clusters into **74 matched pairs**
- Choose clusters within each pair to be as similar as possible
- **Randomly choose one cluster in each pair** for encouragement

Advantages

- Matching controls for **observable confounders**, to a degree
- Randomization controls for observable and **unobservable confounders**, to a degree
- **Pairing provides failure safeguard**: drop entire pair, and treatment and control groups remain equivalent
- One such failure has already occurred

More Detail on Matching Procedure

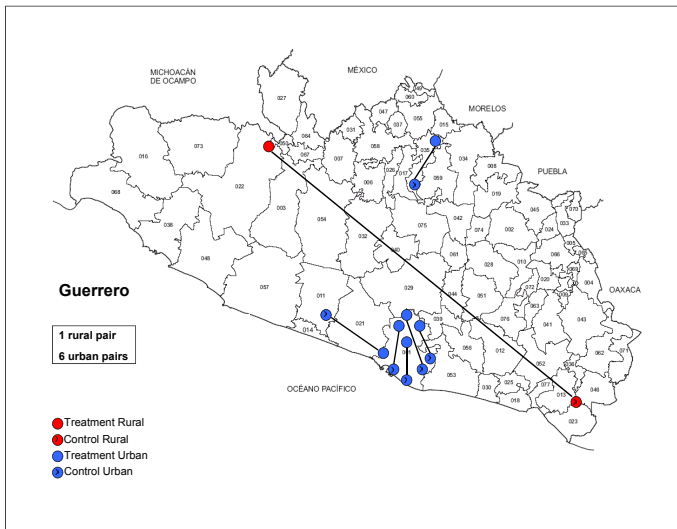
- Select background characteristics
 - Ideally: outcome measures at time 1 (based on a survey done before random assignment)
 - Next best: proxies highly correlated with the outcome measures
 - Practically: All available, plausibly relevant variables (38 covariates for both Rural & Urban; 30 in common)
 - demographic profiles
 - socioeconomic status
 - health facility infrastructure
 - geography and population
- Exact match on state and urban/rural
- Compute “distance” between every possible pair of clusters (using Mahalanobis Distance, normalized with all state-validated clusters)
- An “optimally greedy” matching algorithm:
 - Select matched pair with smallest distance between clusters
 - Repeat until all clusters are used

Experimental Design Implementation

- At the last moment: Flip coin to choose treatment and control cluster for each pair
- Treatment assignments delivered to state governments
- Intensive affiliation begins in treatment clusters
- 74 matched treatment-control pairs in the evaluation: 55 rural and 19 urban in 7 states

State	Rural Pairs	Urban Pairs	Total
Guerrero	1	6	7
Jalisco	0	1	1
México	35	1	36
Morelos	12	9	21
Oaxaca	3	1	4
San Luis Potosí	2	0	2
Sonora	2	1	3
<i>Total</i>	<i>55</i>	<i>19</i>	<i>74</i>

Matched Pairs, Guerrero

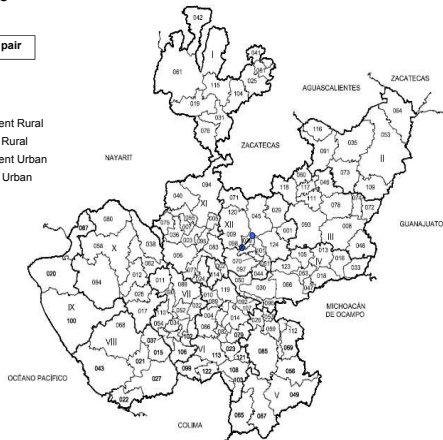


Matched Pairs, Jalisco

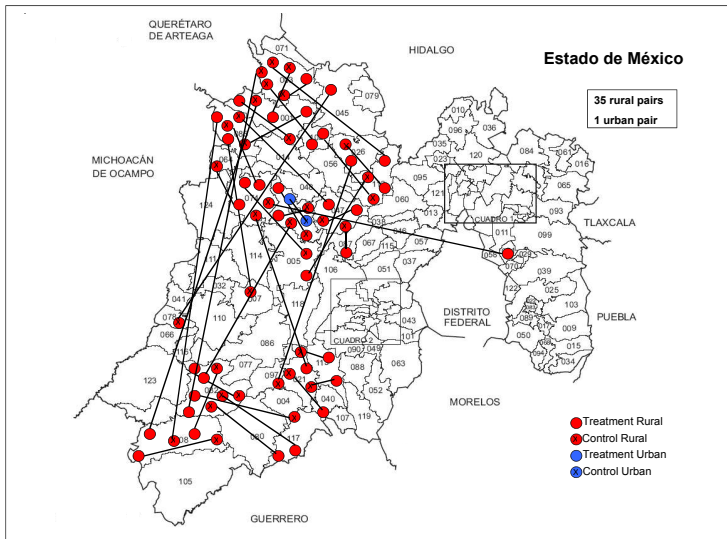
Jalisco

1 urban pair

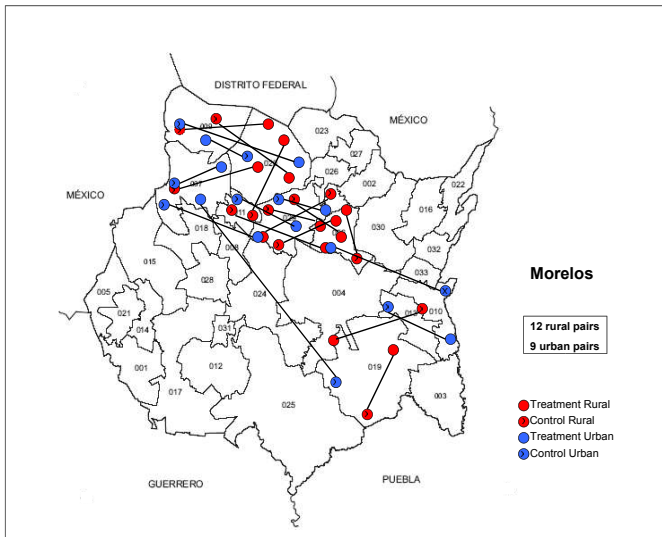
- Treatment Rural
- Control Rural
- Treatment Urban
- Control Urban



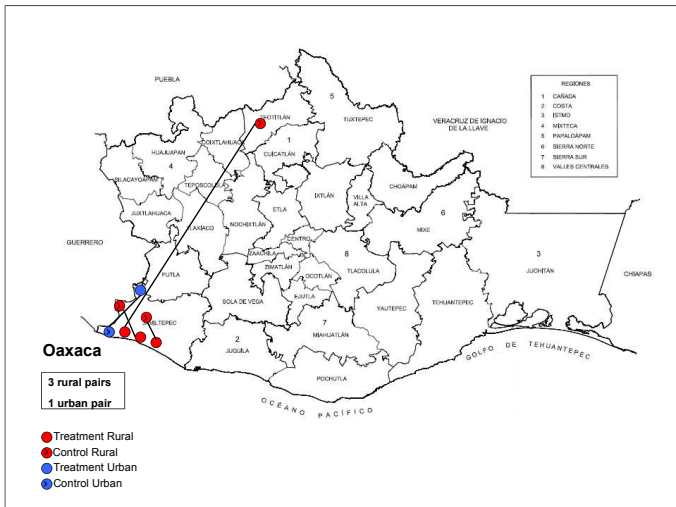
Matched Pairs, Estado de México



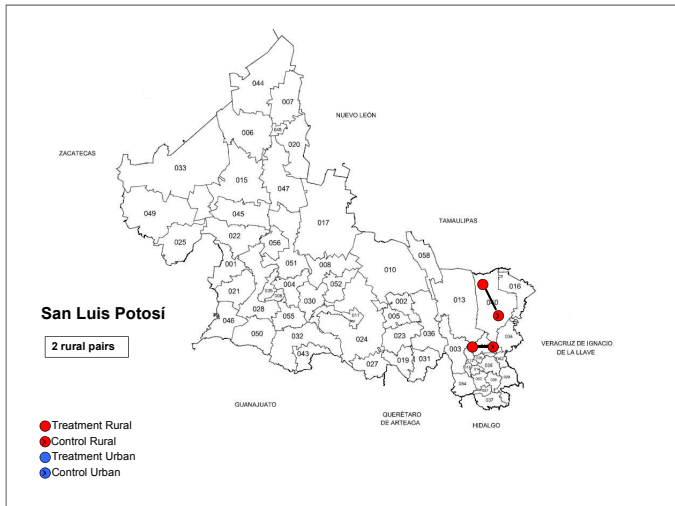
Matched Pairs, Morelos



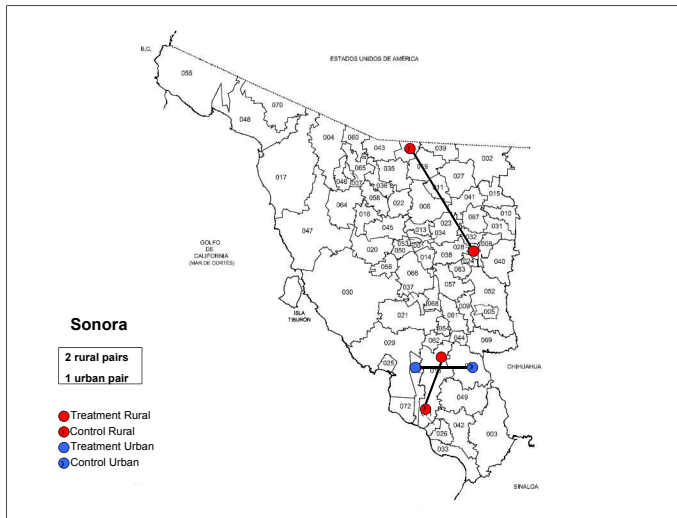
Matched Pairs, Oaxaca



Matched Pairs, San Luis Potosí



Matched Pairs, Sonora



Evaluation Design is Triply Robust

Design has three parts

- 1 Matching pairs on observed covariates
- 2 Randomization of treatment within pairs
- 3 Parametric analysis adjusts for remaining covariate differences

Triple Robustness

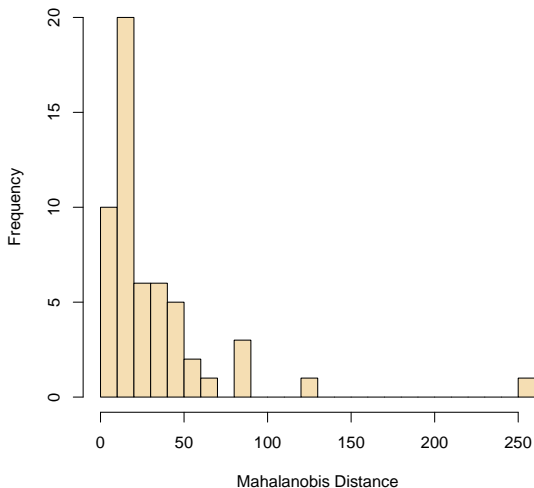
If matching **or** randomization **or** parametric analysis is right, but the other two are wrong, results are still unbiased

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

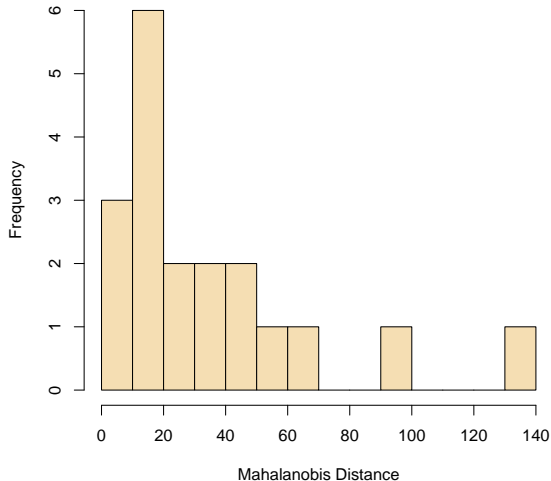
Total Multivariate Distances Within All 55 Rural Pairs

Histogram of Mahalanobis Distances for Rural Pairs, Pre-Assignment



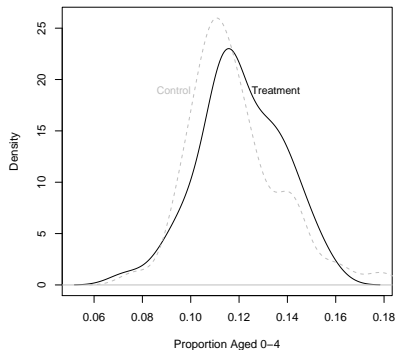
Total Multivariate Distances within All 19 Urban Pairs

Histogram of Mahalanobis Distances for Urban Pairs, Pre-Assignment

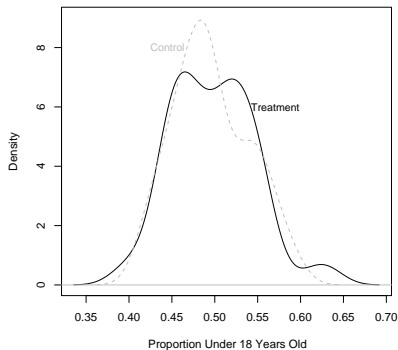


Rural Age Balance After Randomization

Smoothed Histogram of Proportion Aged 0–4, Rural Clusters, Post-Assignment

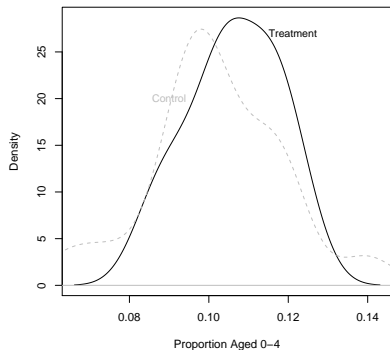


Smoothed Histogram of Proportion Under 18 Years Old, Rural Clusters, Post-Assignment

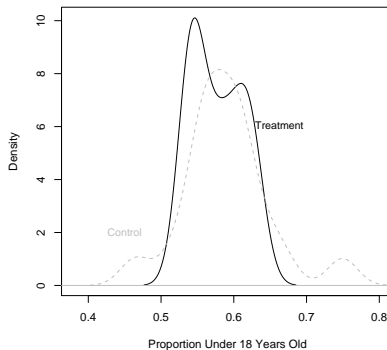


Urban Age Balance After Randomization

Smoothed Histogram of Proportion Aged 0–4, Urban Clusters Post-Assignment

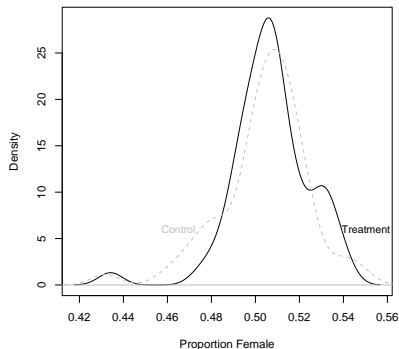


Smoothed Histogram of Proportion Under 18 Years Old, Urban Clusters Post-Assignment

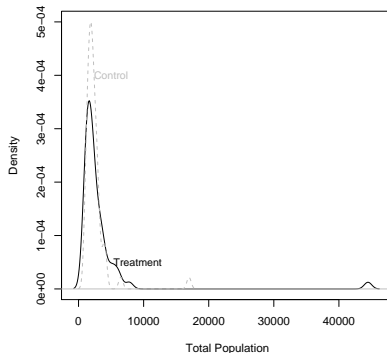


Rural Demographic Balance After Randomization

Smoothed Histogram of Proportion Female, Rural Clusters, Post-Assignment

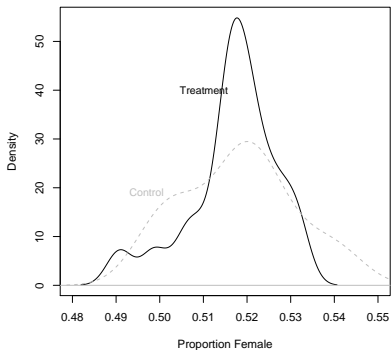


Smoothed Histogram of Total Population, Rural Clusters, Post-Assignment

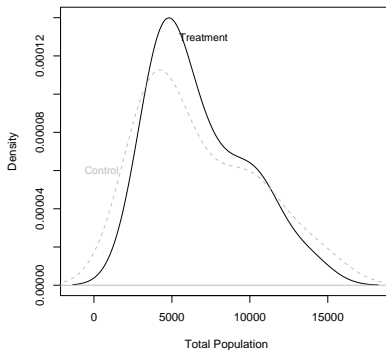


Urban Demographic Balance After Randomization

Smoothed Histogram of Proportion Female, Urban Clusters, Post-Assignment



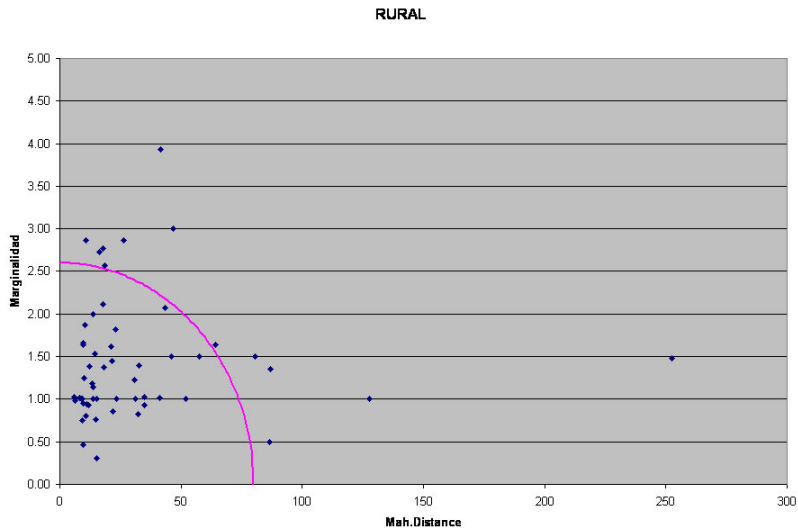
Smoothed Histogram of Total Population, Urban Clusters, Post-Assignment



Household Survey Design

- Baseline in August 2005; followup mid-2006.
- Questionnaire jointly written; implemented by National Institute of Public Health of Mexico (INSP)
- Contents
 - Questions on: expenditure, insurance, Seguro Popular, sociodemographic characteristics, health status, effective coverage, health system responsiveness and utilization, outpatient and inpatient care, social capital, and stress.
 - Physical tests: blood pressure, cholesterol, blood sugar and HbA1c.
- We have 74 matched pairs, but can only (feasibly) survey 50; Sample size: 36,000 households (up to 380 per cluster)
- How to choose?
 - Minimize potential for omitted variable bias by choosing pairs with smallest Mahalanobis Distance
 - Reduce non-compliance problems by including highest percentage of population in incomes in deciles I and II (automatically affiliated)
- Result: 45 rural and 5 urban pairs
- Remaining 24 pairs: also used with aggregate outcomes

Choosing Pairs for the Survey



Health Facilities Survey

- Sample size: 148 health units (corresponding to the pairs of health clusters in the study).
- Panel design
 - first measurement (baseline) in October 2005.
 - follow-up measurement in July-2006.
- Design and implementation:
 - Survey questionnaire designed by Harvard Team
 - Implementation by INSP
- Contents
 - Information on health unit operation, office visits, emergencies, personnel, infrastructure and equipment, and drug inventory.
 - Information on admissions and discharges.

Expert Opinions from Ministry Workshop 4/26/06

(Experts from DGED, INSP, SP, NCGERH, NCESDC)

Dr. Francisco Garrido

Dr. Esteban Puentes

Dr. Gustavo Olaiz

Laura Mendoza

Lic. Ricardo Forero Paez

Dr. Javier Eduardo Figueroa Zuniga

Dr. Miguel Angel Pena Azuara

Dr. Fernando Escalona Figueroa

Dr. Adolfo Valdez Escobedo

Lic. Luis Carlos Fragoso

Ing. Francisco Salcedo

Dr. Emilio Herrera

Dr. Cuitlahuac Ruiz Matus

Dr. Adrian Delgado Cara

Dra. Liliana Martinez Peafiel

Dra. Ana Mara Sols

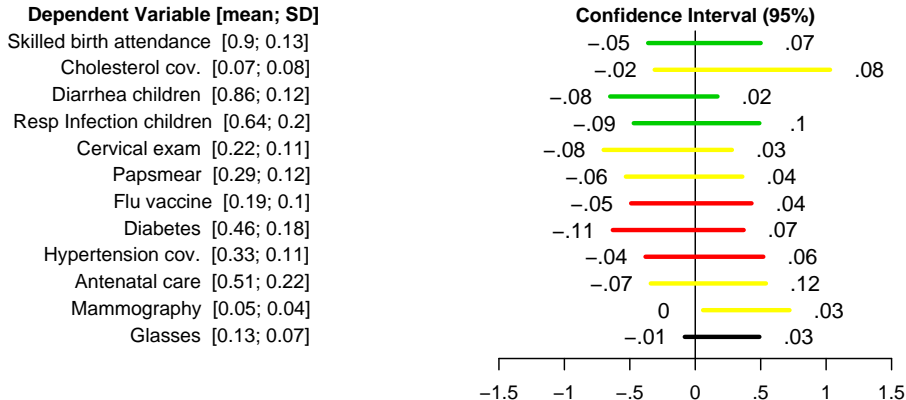
Marisela Cano Bustamante

Dra. Mara Esther Lozano Dvila

Dra. Haidee Caballero Cruz

Effect of SP Rollout at Baseline: 1 of 3

(Expected effects at 10 months: **small**, **medium**, **large**)

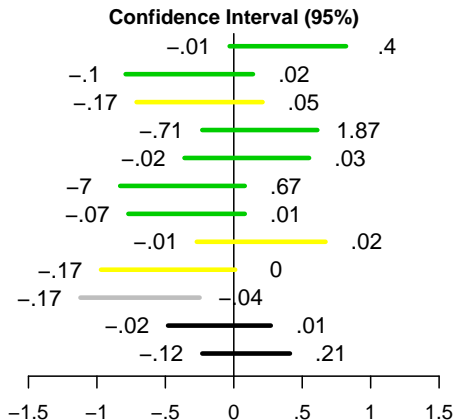


Effect of SP Rollout at Baseline: 2 of 3

(Expected effects at 10 months: **small**, **medium**, **large**)

Dependent Variable [mean; SD]

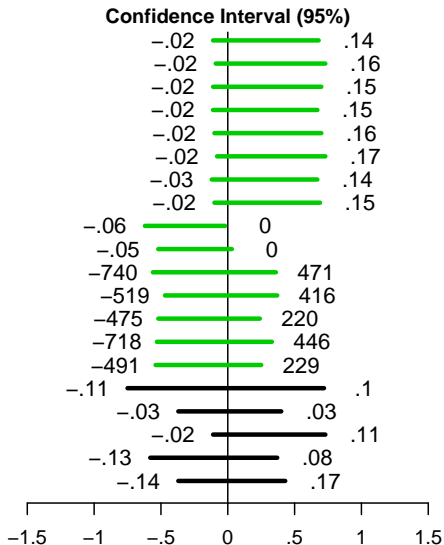
Outpatient visits	[1.24; 0.49]
Prescribed drugs	[1.2; 0.12]
Waiting time	[2.32; 0.23]
SBP	[126; 3.05]
Hypertension	[0.18; 0.05]
Cholesterol	[173; 8.86]
High cholesterol	[0.16; 0.09]
Inpatient visits	[0.09; 0.04]
Cleanliness	[2.04; 0.17]
Talk privately	[2.01; 0.15]
Smoking	[0.11; 0.05]
Seatbelt	[4.75; 0.5]



Effect of SP Rollout at Baseline: 3 of 3

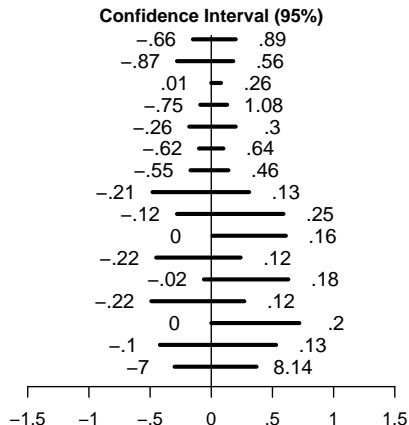
(Expected effects at 10 months: **small**, **medium**, **large**)

Dependent Variable [mean; SD]
Catastrophic (1,30%) [0.48; 0.2]
Catastrophic (1,40%) [0.45; 0.21]
Catastrophic (2,30%) [0.47; 0.21]
Catastrophic (2,40%) [0.45; 0.22]
Catastrophic (3,30%) [0.42; 0.23]
Catastrophic (3,40%) [0.41; 0.23]
Catastrophic (4,30%) [0.45; 0.21]
Catastrophic (4,40%) [0.44; 0.22]
Catastrophic (5,30%) [0.18; 0.1]
Catastrophic (5,40%) [0.16; 0.1]
Out of pocket (1) [3002; 1327]
Out of pocket (2) [2674; 1113]
Out of pocket (3) [1488; 915]
Out of pocket (4) [2320; 1346]
Out of pocket (5) [1488; 915]
Affiliation [0.09; 0.14]
Satisfied health [0.89; 0.08]
Trust local government [0.29; 0.15]
Reduce rich-poor diff. [3.42; 0.21]
Privatize electricity [3.3; 0.39]

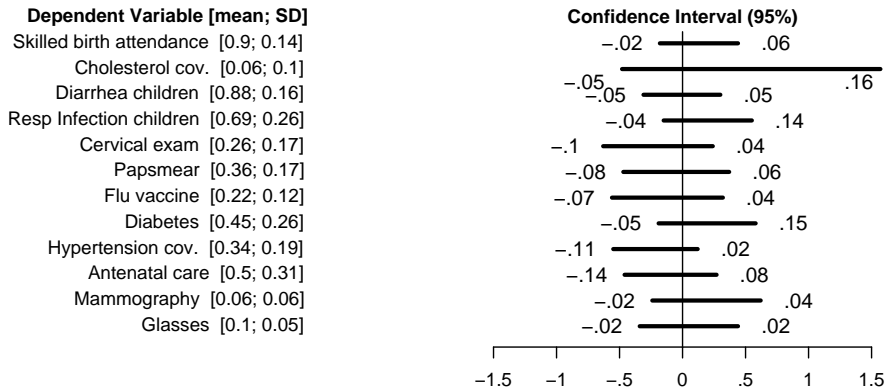


Effect of SP Rollout at Baseline Facilities

Dependent Variable	[mean; SD]
Doctors	[2.69; 4.49]
Family/general doctors	[2.02; 3.04]
Doctors with specialty	[0.55; 3.09]
Nurses	[3.27; 8.38]
Technical personnel	[0.48; 1.47]
Camas censables	[2.47; 6.17]
Camas no censables	[1.18; 3.29]
Stretchers	[0.24; 0.43]
Dental unit	[0.25; 0.43]
Ambulatory surgery room	[0.08; 0.27]
Delivery room	[0.66; 0.48]
Incubators	[0.09; 0.29]
Pharmacy	[0.73; 0.45]
Ambulance	[0.09; 0.28]
Vehicles	[0.06; 0.25]
Weekly hours open	[44; 22]



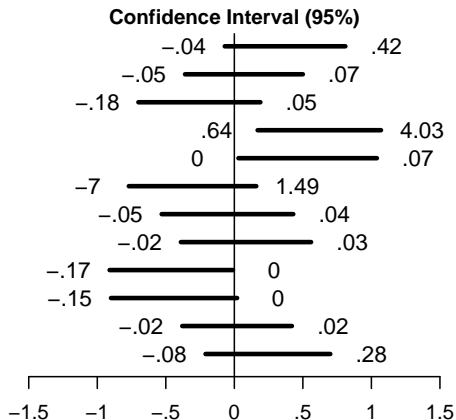
Effect of SP Rollout at Baseline on the Poor: 1 of 3



Effect of SP Rollout at Baseline on the Poor: 2 of 3

Dependent Variable [mean; SD]

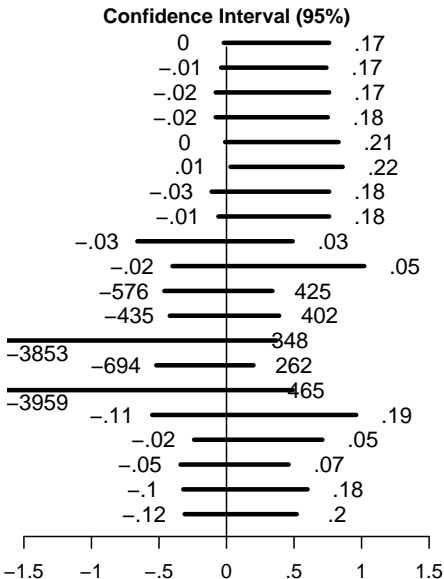
Outpatient visits	[1.29; 0.52]
Prescribed drugs	[1.19; 0.14]
Waiting time	[2.31; 0.25]
SBP	[125; 3.78]
Hypertension	[0.17; 0.06]
Cholesterol	[172; 9.12]
High cholesterol	[0.15; 0.09]
Inpatient visits	[0.09; 0.05]
Cleanliness	[2.04; 0.19]
Talk privately	[2.01; 0.17]
Smoking	[0.11; 0.06]
Seatbelt	[4.97; 0.4]



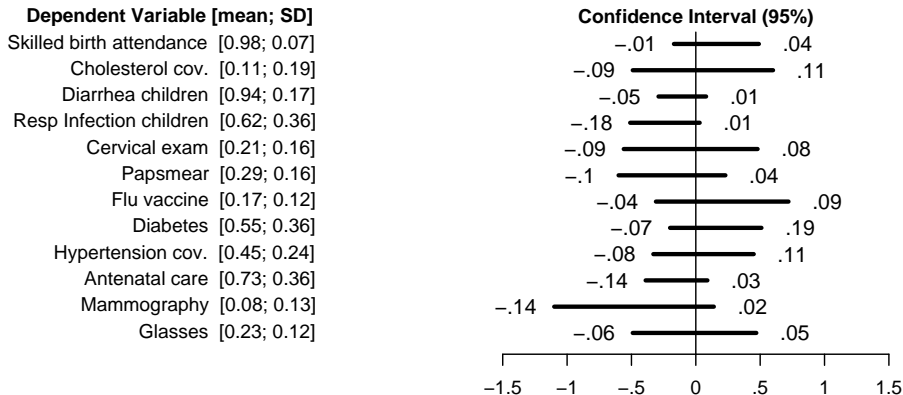
Effect of SP Rollout at Baseline on the Poor: 3 of 3

Dependent Variable [mean; SD]

Catastrophic (1,30%) [0.43; 0.22]
 Catastrophic (1,40%) [0.41; 0.23]
 Catastrophic (2,30%) [0.42; 0.22]
 Catastrophic (2,40%) [0.4; 0.23]
 Catastrophic (3,30%) [0.37; 0.25]
 Catastrophic (3,40%) [0.37; 0.25]
 Catastrophic (4,30%) [0.4; 0.23]
 Catastrophic (4,40%) [0.39; 0.24]
 Catastrophic (5,30%) [0.08; 0.05]
 Catastrophic (5,40%) [0.06; 0.05]
 Out of pocket (1) [2552; 1263]
 Out of pocket (2) [2308; 1035]
 Out of pocket (3) [1316; 951]
 Out of pocket (4) [1931; 1330]
 Out of pocket (5) [1316; 951]
 Affiliation [0.13; 0.2]
 Satisfied health [0.9; 0.07]
 Trust local government [0.3; 0.16]
 Reduce rich-poor diff. [3.39; 0.3]
 Privatize electricity [3.23; 0.39]



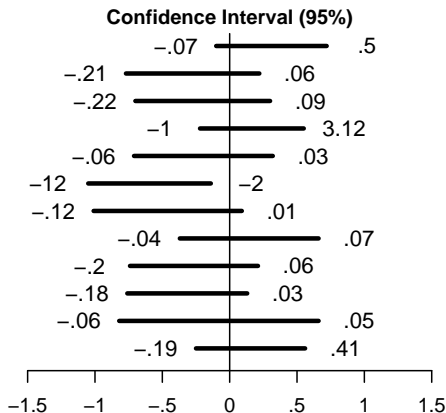
Effect of SP Rollout at Baseline on the Wealthy: 1 of 3



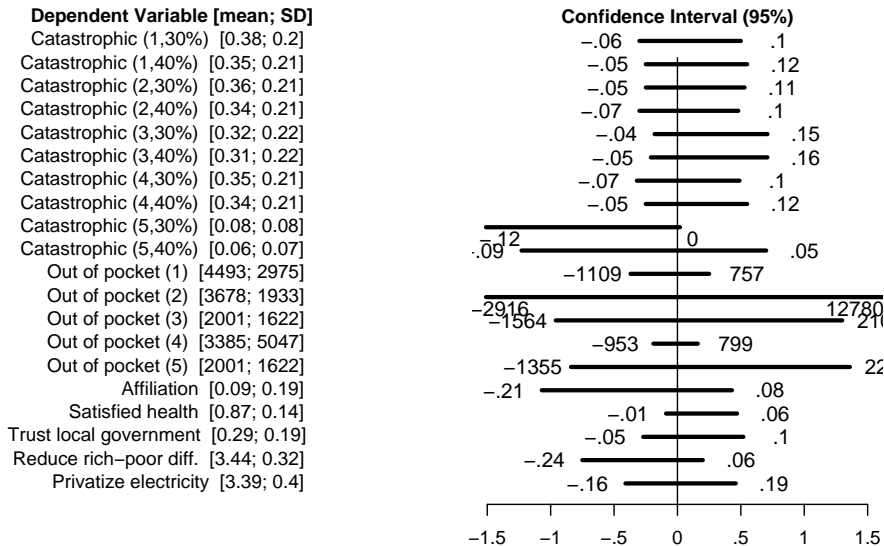
Effect of SP Rollout at Baseline on the Wealthy: 2 of 3

Dependent Variable [mean; SD]

Outpatient visits [1.45; 0.7]
Prescribed drugs [1.18; 0.28]
Waiting time [2.31; 0.31]
SBP [125; 5.69]
Hypertension [0.16; 0.09]
Cholesterol [175; 11]
High cholesterol [0.18; 0.12]
Inpatient visits [0.11; 0.1]
Cleanliness [1.99; 0.27]
Talk privately [2; 0.24]
Smoking [0.11; 0.08]
Seatbelt [4; 0.74]

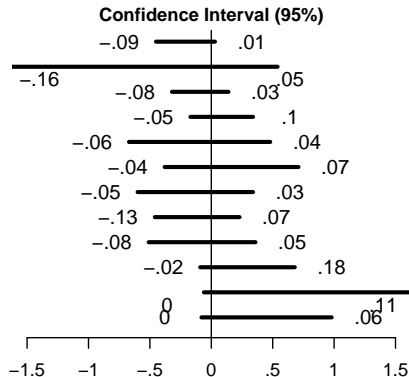


Effect of SP Rollout at Baseline on the Wealthy: 3 of 3



Effect of SP Rollout at Baseline on Others: 1 of 3

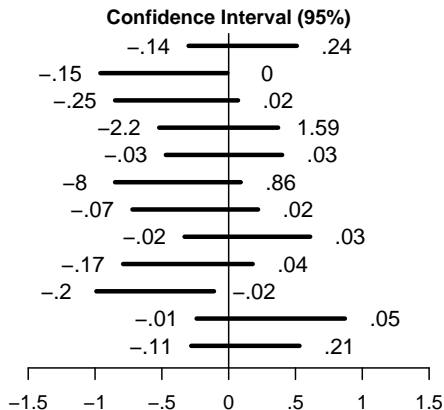
Dependent Variable [mean; SD]
Skilled birth attendance [0.89; 0.2]
Cholesterol cov. [0.06; 0.08]
Diarrhea children [0.8; 0.24]
Resp Infection children [0.63; 0.3]
Cervical exam [0.16; 0.09]
Papsmear [0.21; 0.1]
Flu vaccine [0.14; 0.09]
Diabetes [0.39; 0.28]
Hypertension cov. [0.27; 0.15]
Antenatal care [0.53; 0.27]
Mammography [0.04; 0.04]
Glasses [0.11; 0.06]



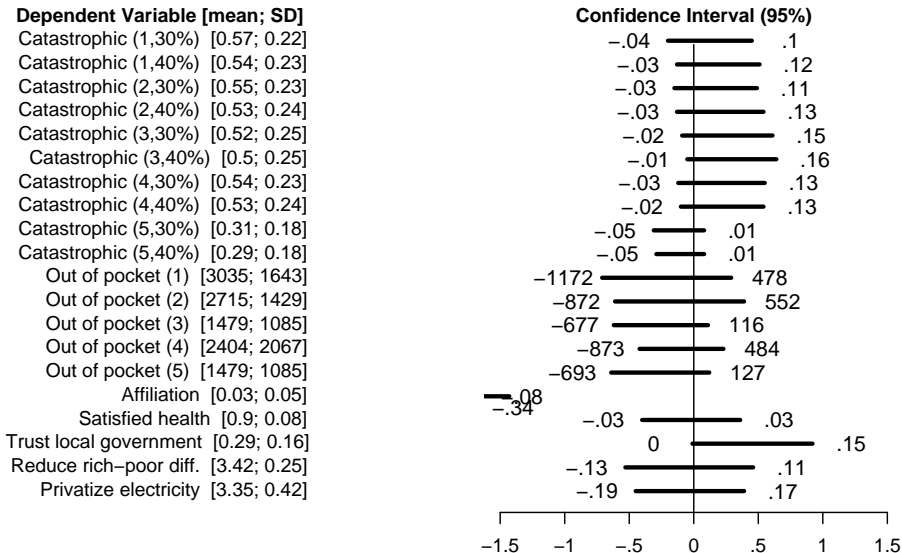
Effect of SP Rollout at Baseline on Others: 2 of 3

Dependent Variable [mean; SD]

Outpatient visits	[1.08; 0.46]
Prescribed drugs	[1.22; 0.16]
Waiting time	[2.31; 0.29]
SBP	[126; 4.26]
Hypertension	[0.19; 0.07]
Cholesterol	[173; 9.58]
High cholesterol	[0.16; 0.1]
Inpatient visits	[0.1; 0.05]
Cleanliness	[2.04; 0.21]
Talk privately	[2.02; 0.21]
Smoking	[0.11; 0.06]
Seatbelt	[4.86; 0.4]



Effect of SP Rollout at Baseline on Others: 3 of 3



For more information

<http://GKing.Harvard.edu>