Major Research Paper

SOCIAL POLICY:

HOW DO WE KNOW WHAT WORKS?

The Case of Progresa

By
Heather Scott
5959885

Professor Miles Corak
University of Ottawa
November 21, 2012

Executive Summary

This paper looks at the use of social experimentation as an evaluation tool to examine social policy. It is often used to examine causal relationships. The paper reviews the theory behind social experiments, and discusses advantages as well as possible limitations. The Progresa-Oportunidades program in Mexico is used as a case study to illustrate the usefulness of experiments as a tool for program evaluation. Progresa illustrates how social experiments work in practice, and the important role evaluations play. In Progresa's case the positive results led to the continuation and expansion of the program after a historic change in government, and it is now Mexico's principal anti-poverty strategy.

Contents

1. Introduction

Evaluation systems examine programs to see if they are accomplishing their objective. Programs devoid of evaluation research can lead to the perpetuation of less effective and more costly policies. Social experiments are a tool for evaluations and have the advantage of showing causal impacts, thereby highlighting the good and exposing the faulty of a policy (Worthen 1997, p. 23). The Treasury Board of Canada defines evaluation as “a systematic collection and analysis of evidence on the outcomes of programs to make judgements about their relevance, performance and alternative ways to deliver them or achieve the same results.” Its objective is to use the evidence for program improvements, expenditure management, cabinet-decision making and public reporting (Treasury Board of Canada Secretariat, 2009). Favouring this definition of evaluation, the paper will discuss social experiments as an evaluation tool, seek to explain how they work in theory, and then illustrate how they work in practice using the Progresa-Oportunidades program in Mexico as an example. Policies aimed at alleviating poverty are of particular interest as NGOs, private companies and governments spend enormous amounts of resources in this area. For this reason it is worthwhile knowing if the various policies employed have met their objectives.

When social experiments are used as an evaluation tool they show causal relationships. This is possible because of the use of randomized control trials. Randomization overcomes the problem that we cannot simultaneously observe an individual in a treated and untreated state. It does this by randomly allocating members of a target population into treatment and control groups. While both groups, on average, have similar characteristics one group receives the treatment while the other group doesn’t. Results from the experiment show the averages of the groups. The difference in these averages is used to measure the effect of the treatment. This information makes a significant contribution to the evidence available to policymakers, for making decisions about policy.

The paper has five sections and is organized as follows. Section 2 looks at the theory behind social experiments, and the discussion includes advantages of social experimentation as an evaluation tool as well as potential threats it faces. Possible disadvantages and limitations are also discussed. Section 3 introduces the Progresa-Oportunidades program in Mexico as a case study, giving a background and description of the program's features. Section 4 illustrates how the design Progresa-Oportunidades included a strategy for its evaluation from the first stage of its implementation, and section 5 presents the results of that evaluation. Although Progresa produced a number of interesting outcomes, this section will focus on the validity of the randomization process and health. Section 6 concludes the paper.

Experiments have advantages and limitations. The review of Progresa highlights the advantages of random assignment in addressing selection bias and the ease with which results can be communicated. Practitioners need to be aware that when discrepancies between the theoretical capabilities of an experiment and the running of it appear, there is a threat to its internal design. Progresa was able to minimise the limitations of experimentation through good planning and implementation, and capitalising on the support it had from the highest management in the country.

2. Theory Behind Social Experiments

Both experimental and nonexperimental methods of evaluation face the common problem of establishing a credible basis of comparison because no person is observed simultaneously in both the treated and non treated state. In 1935 R.A. Fisher proposed that the only way of achieving unbiased results was to assign subjects to treatment "wholly at random" (Burtless

1995, p.66). This kind of study is what is referred to as a randomized or social experiment. It is different from other studies in that it uses random assignment in selecting participants to receive different treatments.

Social experiments are one of the clearest methods for inferring causal relationships. Individuals cannot be observed in both the treated and untreated states and so participants in a study are assigned to two different groups. Random assignment of participants to these groups means that the groups, on average, are similar in terms of observable and unobservable characteristics. One group is the treatment group, which receives a type of treatment, while the other group is the counterfactual. In this group, known as the control group, treatment is withheld from members. Causal effects are established by comparing the outcomes of the treatment and control groups. Heckman and Smith (1995), in their article on "Assessing the Case for Social Experiments", argue that the credibility of both experiments and nonexperiments rest on two critical factors. First the reliability of data used to generate answers, and second the widespread acceptance of two crucial assumptions: randomization does not alter the pool of participants or their behaviour, and no alternative sources for similar treatment can be found. The following section analyses the advantages and disadvantages of using randomization in experiments in light of these factors, drawing on the work of Heckman and Smith (1995) and Burtless (1995).

2.1 Randomization in Social Experiments

2.1.1 Advantages

Selection Bias

Random assignment ensures each participant in a study has the same opportunity to be assigned to either the treatment or control group. An example of how randomization works in practice can be drawn from a ball machine, like those seen on television game shows. Imagine the machine is full of balls, 20% are red and 80% are blue. A handle is attached to a turntable which provides momentum inside the machine to push the balls around, until one is thrust through a small opening. Accordingly participation in a treatment group depends on which colour ball you draw, red means you're in the treatment group and blue means you're not. Thus selection is independent of all other factors, whether observed or unobserved. Similarly, the manager of an experiment should use an unbiased process that randomly selects participants for their study, and not rely on their own decisions about who should be in what group.

The most distinguishing characteristic of using random assignment in allocating participants of an experiment to treatment and control groups is that the procedure effectively addresses selection bias. Selection bias arises from using a nonrandomly selected target population to estimate behavioral relationships (Heckman, 1979). This happens as a result of two factors: first individuals select themselves to participate as a subject in an experiment, or second the selection of samples is decided by the researcher to support a particular hypothesis. This means program participants differ from non-participants in unobserved characteristics which affect both the person's decision to participate and the outcome of the program. Participants may have the most to gain from a particular treatment and are more motivated to commit to program activities. Inevitably the outcomes reflect the impact of the program on motivated participants but not how the program on average would affect the target population (World Bank, 2012). For example, it is possible that a smoking cessation program would attract individuals who are motivated to participate and thus prepared to commit to the program's activities. The motivation of a person who wants to stop smoking in order to be eligible for a lung transplant is different from the average smoker. If the person intentionally selects them-self to be part of a treatment group in a smoking cessation program, self-selection has occurred and the sample is unrepresentative of the population of people who smoke. The outcome of the program is based on biased information and does not reflect the average effect of the program on people who smoke. The underlying rationale for randomization is that every member of a target population has an equal, nonzero chance of being chosen for the treatment group. It is not a choice-based selection but follows strict procedures so that every person has an equal chance of being in the treatment group. When these conditions are met it is highly likely that participants selected into the treatment and participants not selected into the treatment will be alike on all observed and unobserved dimensions (Weiss 1998). It is fundamental to the internal validity of the experiment to get individuals of similar motivation and then separate them into two groups. When randomization is not possible, observational studies can be used to observe the effect of a treatment. The evaluator can use a number of observed characteristics as controls to condition the dependent variable (there is no independent variable in this kind of study). However when unobserved characteristics are present, inferences about the outcomes of the effect of a treatment are based on bias information. Observing the effects of a smoking cessation program may control for observable characteristics such as income and gender but does not address the role played by unobservable characteristics such as motivation or ability.

Once participants have enrolled in a program, randomization addresses selection bias on the basis of two assumptions:

1) It does not alter the pool of participants or their behaviour. This means that those who are allocated to the treatment group do not differ from those who would have been allocated to the control group. No in-built bias has taken place.
2) Close substitutes for the treatment are not readily available. If there were and participants in the control group opted out of the experiment to seek them, the control group would no longer correspond to the desired untreated state of persons who

wanted to receive treatment but did not. The outcomes would no longer provide an estimate of the mean impact of treatment on the treated.

Given these assumptions, the results which follow from a randomized experiment can therefore be an internally valid estimate of the average treatment effect.[1]

Communication of Results

Burtless (1995), highlights a further advantage of social experiments. That is the ease with which the findings can be communicated. He found that recent experiments which are narrow in their focus have become more numerous and have had a larger impact on actual policy decisions. He attributes this to the relative simplicity of social experiments which allows the outcomes to be communicated in straightforward language; unlike the findings of nonexperiments which are subjected to a variety of qualifications. As a consequence, not only are results acceptable to social scientists but are also accessible to individuals with no formal training in economics or statistics. These may include policymakers. The benefit to this is an increased chance of broader political acceptance of the findings which implies that results are used for changing public policy.

2.1.2 Potential Threats to Experimental Design

Sampling Error

An internal threat to the design of the experiment is allowing errors to occur in the process of gathering the sample. These may be inadvertent or more purposeful (self-selection). Managers running the experiment, for various reasons, could choose favourable participants for the treatment group thus incurring the error of selection bias. In theory the researcher has control in assigning study participants, however in practice experiments sometimes rely on those who

[1] An "internally valid" estimate is an unbiased measure of the treatment effect in the sample enrolled in the experiment.

implement programs to randomize the treatment and so the researcher may move away from influencing the intervention. To maintain the integrity of the design, management must establish and constantly supervise procedures to ensure the quality of the data being collected. Good planning and implementation can achieve internal validity. In selecting the team to run the experiment the following should be taken into consideration:

1) Confidence in the professional skills of the evaluators
2) Evaluators who have no stake in the program so they are insulated from biasing their data or interpretations and are seen by others as impartial
3) They must understand the program and have appropriate training
4) Have autonomy to introduce alternatives when challenges arise

Theory versus Practice

Heckman and Smith (1995, p.107) argue that the advantage of randomization in social experiments must not be overstated as there is a sizeable divergence between theoretical capabilities of random assignment evaluation and the practical results. Despite putting measures in place prior to the experiment, the practical running of an experiment can risk internal threats to its design. Relying on the training of staff at various experimental sites is no guarantee that the process of randomization is undertaken successfully. Administrators may switch people in groups due to time constraints or unavailability of participants or they could have a stake in making the program look good. Participants may also, for various reasons, refuse to participate in one of the groups thereby causing the treatment and control group to no longer be similar. It has been observed that the later randomization occurs in the selection process the less likely are people to refuse to participate (Weiss 1998). Sample attrition can also jeopardize the experiment as members of the sample may be lost over time. This causes changes in the two groups that were randomly selected at the beginning of the experiment and are now unrepresentative of the population originally enrolled. Measures can be taken to counter attrition, by keeping in touch with participants at frequent intervals or offering incentives. Though these measures may not eliminate the problem of nonresponse entirely, one way of avoiding it would be to rely on administrative data such as program benefit records, unemployment insurance wage records, or social security earnings records (Burtless and Orr 1986). These potential threats lead us to discuss limitations of random assignment in experiments.

2.1.3 Disadvantages and Limitations

Institutional factors

Heckman and Smith draw attention to the limitations placed on the use of randomized experiments due to the institutional structure of social programs.

1) The optimum time for random assignment would be at the initiation of treatment so as to minimise the problem of attrition within the treatment group, but this is not always possible because of political and institutional factors. Researchers face practical considerations which limit their ability to carry out random assignment at the optimum place. In practice this may mean a delay of weeks or months, for instance if treatment involves having to rely on an academic schedule for a new semester to start, participants may have moved out of town, lost interest or found an alternative service.
2) Institutional factors can make it impossible to produce separate experimental estimates of the impact of different service types. For instance an individual may be receiving two or more services but only end up using one. It is unknown which set of services a client will definitely receive. Multistage randomization will need to be used in order to obtain experimental estimates of the impacts individual services. This is costly, difficult to implement, disruptive and a large sample size is needed.

Partial Equilibrium Results

Burtless (1995) raises the problem of partial equilibrium results. Findings can be criticised for not reflecting the general equilibrium effect of a service treatment. If the experiment is on a small-scale then the benefits of the treatment are conferred on only a small fraction of the people who would benefit under a full-blown national program. This does not recognise how the behaviour of those receiving the treatment may influence the rest of the population. This is particularly the case if a large fraction of the population is eligible for the program. This has consequences for the external validity[2] of the experiment. Nonetheless it may be possible to predict the general equilibrium effect of a fully-fledged program even if it cannot be determined.

Cost

Random Control Trials may have a significant cost. A great deal of real resources is needed. Nonexperimental evaluations using existing data can be much less expensive. The experiment may span over several years as it goes through the process of design, implementation and finally analysis, which, when a policy decision about a particular policy cannot be deferred, would be less useful. By the time results are available the significance of the issue may have faded. On the other hand the issues looked at are still central to policy making even if the specific policy is no longer being considered, and thus have value (Burtless and Orr 1986). Many issues central to effective policy design do not change. As it is costly to develop, implement and administer a new treatment obtaining funding is more difficult to secure than analysing existing information about past conditions. What should be borne in mind however is that programs devoid of evaluation research can lead to the perpetuation of less effective and more costly

[2] External validity" refers to the extent to which research findings can be generalized to the larger population and applied to different settings.

policies than alternatives, and as Burtless and Orr (1986) argue, the value of an experiment is greatest when the tested treatment is found not to be cost effective.

From an Ethical View

There is a school of thought which argues that random assignment in experiments is an unethical way to ration public resources. It reasons that it is unethical to deny benefits to eligible members of the population for the purposes of a study. Within the context of an ongoing program, withholding potentially beneficial services from the control group would make them worse off. However this presumes the treatment works. From a researcher's point of view nobody is being denied anything because it is not known whether the treatment works or not. Burtless and Orr (1986) argue from the premise that the fundamental ethical principle should be that the expected net benefits to participants of both the treatment and control groups should not be less than those they could expect in the absence of the experiment. In an experiment which provides a beneficial treatment that would otherwise not have been available, this condition is met. The treatment group is better off because of their participation, and the control group is no worse off because of their participation. No harm has been done. Furthermore there are circumstances where it is fitting to deny services to a control group. Every experiment takes place within the parameter of a budget and therefore it may not be possible for the treatment service to be available to all eligible applicants. In this respect random assignment would be a fair way to select the individuals when resources are limited. Everyone would have the same, nonzero chance of being chosen.

Ethical regulations which require each participant to view the experiment as being beneficial to them can be met by obtaining informed consent. This is usually the case for experiments that provide treatments which would otherwise not be available. When participants agree to the conditions of the experiment they are in fact acknowledging that their expectation for the experiment is to yield positive benefits. Those individuals who refuse to consent must then receive the same services they would have in the absence of the experiment. Good experimental design would obtain informed consent of participants, and have in place confidentiality measures to ensure privacy is respected. It may also be necessary to offer compensation to participants in order to meet the research objectives as well as satisfy ethical requirements, when too few applicants voluntarily choose to participate.

To conclude, random assignment does not prevent researchers from using nonexperimental methods to analyse data, and it remains a powerful source of identifying information to measure the average effects of the treatment offered. Much policy making happens based on the average effects.

3. The Progresa-Oportunidades Program in Mexico

Programa de Educación, Salud y Alimentación (the Education, Health, and Nutrition Program), known by its Spanish acronym, PROGRESA, was launched in 1997 to combat poverty. Known as Oportunidades since 2002,[3] its goal was to enhance the human capital of those living in extreme poverty. Being a "homegrown" initiative, the initial work of Progresa was done by a small group of social scientists[4] headed by Santiago Levy, then Deputy of Finance. Levy, who holds a doctorate in Economics, is credited as being the architect of the program and was responsible for its initial implementation. A distinguishing feature of the program is its integrated approach to poverty relief. Although it is known that political, economic and social constraints (reflected in the fixed and limited budget for Progresa), mean that the

[3] Throughout this paper the program is referred to as Progresa.

[4] This group consisted of Jose "Pepe" Gómez de León (a demographer who headed CONAPO, Mexico's National Commission on Population and was the first National Coordinator of PROGRESA), Evelyne Rodriguez (who worked with Levy in the Finance Ministry) and Daniel Hernández (who worked with Gómez de León at CONAPO, was the first Assistant National Coordinator of PROGRESA and succeeded Gómez de León as National Coordinator of PROGRESA), (Behrman, 2007, p.2).

program is not sufficient to eliminate poverty completely (Skoufias, Davis and Behrman 1999), it has had a considerable impact on the lives of the poor and has been praised in articles by leading economists like Gary Becker (1999), a Nobel Laureate in Economics, Alan Krueger (2002) a prominent economist at Princeton University, and the well-known American political scientist and political economist Francis Fukuyama (2007). Aspects of the program have been imitated in a number of developing countries in Latin America, Africa and Asia and in developed countries, for example Opportunity NYC is a program in New York City modeled on Progresa (Behrman 2007). Introduced initially in small, rural communities, by the end of 2005 the program had expanded to about 5 million families, representing practically all households living in extreme poverty (Levy 2006). It is Mexico's largest, single poverty alleviation program in history.

3.1 Background

The political and economic climate against which the program was developed is of as much importance as the design and implementation of Progresa-Oportunidades. The years 1994-1995 were ones of political uncertainty and economic crisis. Against a backdrop of political unrest in the state of Chiapas and the murder of a presidential candidate, the government transitioned from President Carlos Salinas to President Ernesto Zedillo. In December 1994 Mexico suffered a major macroeconomic crisis which resulted in the largest fall in economic activity in more than 50 years, over the course of 1995 the GDP decreased by 6% (Levy 2006, p. 13). Concerns focused on the implications of the shock for the poor, and debate was centred on existing programs for the poor: should they be strengthened or were they adequate?

Poverty in Mexico is measured by the National Council for Evaluation of Social Development Policy. Their methodology combines three aspects of life conditions:

socioeconomic wellbeing, social rights and territorial context. They published the following definitions for poverty and extreme poverty (Oportunidades 2012a). Someone is considered to be in poverty when their income is not enough to satisfy their basic needs and present inadequacies in at least one of the following indicators:

1. Average gaps in education
2. Access to health services
3. Access to social security
4. Household size and overall housing quality
5. Access to basic household services
6. Access to nutrition
7. Social cohesion.

People are considered to be in extreme poverty when they face three or more insufficiencies each and their full income is not enough to satisfy their basic needs. Extreme poverty is concentrated in rural areas. By mid-1990 an estimated 24% of all Mexican households lived in extreme poverty and more than 50% of all households in rural areas were considered poor (Levy 2006).[5] Fifteen food subsidy programs, including in-kind provision, were in place with eleven of those targeted at the urban and rural population. Many developing countries had adopted these as social assistance programs but they were proving to be expensive and not effective because they were not helping the poor to break out of poverty. For example, the Subsidio a la Tortilla program was a generalized food subsidy program providing price controls to all tortilla buyers. The result was a negligible effect on the poor because of the leakage of benefits to non-poor households, and was costing more than half the total budget for food subsidy programs. It also illustrated the imbalance of funds between the rural and urban areas. Even though 75% of the budget was being spent in urban areas, less than 40% of the poor lived there (Levy 2006, p.5). In kind provision of food baskets and school breakfasts to rural

[5] Non-income indicators of welfare were used to measure this like nutritional status, infant mortality, illiteracy rates, access to and use of health and education services.

populations was problematic because of widespread population dispersion. Transport was difficult and expensive because of mountainous terrain and lack of roads. There was a low volume of people per locality and few storage facilities. Little co-ordination between agencies engaged in the programs often led to duplication of efforts, and failure to identify poor households. Furthermore food subsidy programs failed to support the most vulnerable members of a family, pregnant or lactating women and children under two years old. In addition there wasn't any system for evaluating the programs which allowed for little accountability.

Of major importance in the design of Progresa was the large volume of work done by academics such as economists, sociologists and nutritionists in the 1980s and 1990s. Their work brought advances in understanding poverty and contributed to a better appreciation of the contributing factors of poverty. Stress was placed on the links between food intake, nutrition, health and education, and it was emphasised how an integrated approach to delivering these services to the poor would be more effective. Students learn better on a full stomach because they can be more focused in the classroom, more education leads to better health by being more aware of preventative health measures and so on.[6] The suggestion from this literature is that an integrated approach to poverty alleviation will be more effective and efficient. Some of the following ideas are examples of what is included in an integrated approach:

1) in addition to food subsidies more income is provided to buy food and other necessities
2) information on hygiene and reproductive health is distributed
3) alternate sources of income are provided to replace the income earned by children, so they can attend school more regularly and for longer
4) health risks are modified by more frequent contact with health service providers

Another body of work highlighted the need to involve poor families in overcoming their poverty by giving them greater control of their resources and more choice in spending. Overall

[6] For a more detailed discussion of how these complement one another see Levy, 2006, pp.11-12.

this encourages those involved to take more responsibility, and is key to avoiding a dependence on income transfers.

The economic crisis provided the opportunity for change. President Zedillo, an economist and former minister of budgeting and planning in the previous administration, gave the leadership and support necessary for this change. In stark contrast to previous programs operating in isolation in each of the areas of health, nutrition and education, Progresa was designed to have an integrated intervention thereby addressing all dimensions of human capital simultaneously. In the context of rapid democratization and a difficult budgetary situation, evaluations were made part of its design to enforce transparency and ensure effectiveness.

Progresa was designed as a targeted program so that its resources benefited households living in extreme poverty. Those households who became beneficiaries of Progresa were to stop receiving benefits from other pre-existing programs. In the short run the objective was to minimise duplication of benefits to poor families, while the long run objective was to absorb a variety of poverty alleviation programs within Progresa.

3.2 Program Design

Program design, among other factors, took into account annual budget allocations and whether communities had the necessary education and health facilities. Communities in rural areas were selected using a marginality index calculated from data taken from a national-expenditure survey and a census. It was based on the predicted probability of being poor for each community (Davis 2003 p.29). Within the selected communities, potential eligible households were identified using socio-economic data collected from the households. Benefits were targeted directly to people living in extreme poverty in rural areas through cash transfers, which were conditional on regular school attendance by children of the household, and mandatory visits to health care centres. Cash transfers were given to mothers in the household because they were perceived to use household resources efficiently and effectively, reflecting the immediate needs of the family (Skoufias 2005).

Education Component

Education is considered a strategic factor in breaking the cycle of poverty thus benefits of Progresa included education grants and monetary support to improve school enrollment and attendance. Families needed to be compensated for the income a child could be earning by working instead of attending school. Moreover beneficiaries were not in a position to pay for school even if tuition was free. Education grants were designed so that the size of the grant increased through grade levels. Grants were higher for females because enrollment rates for females in secondary school decreased relative to those for boys. They were adjusted every six months in order to compensate for inflation. All students finishing high school received a one-time cash transfer, and subsidies are given for school supplies. In order for the beneficiaries to receive these grants they needed to attend school for more than 85% of school days each month, from the third grade to the end of high school. Failure to do this would result in a temporary loss of benefits with the possibility of the loss being permanent. Records of attendance are kept by the school. The amount spent on education transfers increased from 131.5 million pesos in 1997 to 2, 797.8 million pesos in 1999 (Appendix 1). This represents an over 20-fold increase in spending.

Table 1.1: *Progresa monthly cash transfer schedule, pesos.*

Grant	*January-June 1998*	*July-December 1998*	*January-June 1999*	*July-December 1999*
Educational grant per child (conditional on school enrollment and regular attendance, in pesos)				
Primary				
3rd Grade	65	70	75	80
4th Grade	75	80	90	95
5th Grade	95	100	115	125
6th Grade	130	135	150	165
Secondary				
1st – male	190	200	220	240
2nd – male	200	210	235	250
3rd – male	210	220	245	265
1st – female	200	210	235	250
2nd – female	229	235	260	280
3rd – female	240	255	285	305
Grant for school materials per child				
Primary - September	-	In-kind	-	110
Primary - January	40	-	45	-
Secondary - September	-	170	-	205
Grant for consumption of food per household (conditional on attending scheduled visits to health centres)				
Cash transfer	95	100	115	125
Maximum grant per household	585	625	695	750

The average exchange rate from January 1998 – December 1999 fluctuated around 0.11-0.12 MXN: 1 USD.
Adapted from Hernandez, Gomez de Leon, and Vasquez (1999) in Skoufias 2005, p.4

Health and Nutrition Component

In an effort to break the intergenerational transmission of extreme poverty, health and nutrition of children in their formative years are important areas of concern as these can improve their physical and cognitive development substantially. Thus Progresa aimed at improving the health of children at conception (Gertler 2000.). It used two strategies to do this. First it provided an income transfer to relax financial constraints (in buying food) and provided medical and nutrition services directly. These strategies complemented one another as receiving the transfers was conditional on participating in the health care program (Gertler 2000, p.3). Health benefits are delivered as a basic package of primary health care services, with a particular emphasis on preventive health care. These services are provided by public health institutions.

Table 1.2: *Composition of the basic health services package*

1. Basic sanitation at the family level
2. Family planning
3. Prenatal, childbirth, and postnatal care of mothers
4. Supervision of nutrition and children's growth
5. Vaccinations
6. Prevention and treatment of outbreaks of diarrhea in the home
7. Anti-parasite treatment
8. Prevention and treatment of respiratory infections
9. Prevention and control of tuberculosis
10. Prevention and control of high blood pressure and diabetes mellitus
11. Accident prevention and first aid for injuries
12. Community training for health care self-help

Adapted from Skoufias 2005, p.5

As nutrition is a crucial determinant of their future development, nutritional supplements are provided to infants, small children and pregnant and lactating women to prevent malnutrition. Mothers are expected to visit a clinic at least once a month to pick up supplies for each targeted household member. The supplements have a shelf-life of about one year and contain specific macro and micronutrients to meet the nutritional needs of mothers and children.

Progresa requires beneficiaries to visit health centres on a regular basis (see Table 1.3) which is done through the setting up and monitoring of a schedule of appointments. Beneficiaries are also expected to attend various health and nutrition talks. The clinic keeps records of attendance and it is their attendance which produces the receipt of cash transfers. While the focus on compliance is of women and children, most household members schedule their annual visit the day of their registration at the clinic. The increase in the amount of health component transfers was modest compared to that spent on education transfers. In 1997 112.9 million pesos was spent compared to 1,305.8 million pesos in 1999 (Appendix 1).

Table 1.3: *Annual frequency of health care visits required by Progresa*

Age Group	***Frequency of check-ups***
Children	
Younger than 4 months	3 check-ups: 7 and 28 days, and 2 months
4 to 24 months	8 check-ups: 4, 6, 9, 12, 15, 18, 21 and 24 months with 1 additional monthly weight and height check up
2 – 4 years	3 check-ups a year: 1 every 4 months
5 – 16 years	2 check-ups a year: 1 every 6 months
Women	
Pregnant	5 check-ups: prenatal period
Postnatal period and during lactation	2 check-ups: in immediate postnatal and 1 during lactation
Adults and youths	
17- 60 years	1 check-up per year
Over 60 years	1 check-up per year

Adapted from Skoufias 2005, p.6

3.3 Scope of Program and Time Patterns

There was a planned strategy for growth of the program across communities and time as the vision for Progresa was that it would eventually cover the entire nation. Progresa was to operate instead of and not in addition to existing food subsidy programs. This plan involved annual budget allocations and logistical considerations connected to running the program in small and remote rural communities. For example, the area had to have educational and health facilities needed to meet the objectives of the program. Most of the discussion about the time patterns of Progresa is drawn from the work of Levy (2006) and Skoufias (2004).

From August 1997 to December 2005 the program expanded rapidly in terms of household coverage, geographical reach, extended benefits and budgetary resources. Not all the benefits in Progresa were there from the beginning. As it has grown in scope the content has been modified, as shown in Table 1.4.

Table 1.4: *Scope of Progresa Benefits, 1997-2005*

Benefit	*July 1997- December 2000*	*January 2001- December 2002*	*January 2003- December 2005*
Nutrition	Nutritional supplements	Same	Improved formulae for supplements
Health Care	Preventive health interventions Health talks	Expanded health topics for mothers and youngsters	Same
Education	School supplies Education grants until secondary school	Education grants until secondary school	Additional cash transfers on completion of secondary school

Source: Levy 2004, p.30.

The expansion strategy gave preference to rural areas because of the greater number of people in poverty and the depth and severity of their poverty. Within rural areas marginal communities were targeted while large urban areas were left to be covered last. The total amount spent of transfers in 1997 was 341.6 million pesos while in 1999 it had grown to 7,089.3 million pesos (Appendix 1). The program grew from covering 300 000 families in 1997 to 6.5 million in 2005 while figures for 2012 measure coverage of 6.5 million families; that means 3 out of 10 Mexican are beneficiaries (Oportunidades 2012b).

__Table 1.5:__ Scope of Progresa Benefits, 1997-2005

Number covered									
Participants	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*
Beneficiary families (thousands)	300.7	1595.6	2306.3	2 476.4	3116	4240	4184.4	5000	5000
States	12	30	31	31	31	31	31	31	31
Municipalities	357	1750	2155	2166	2310	2354	2360	2429	2435
Localities	6344	40 711	53 152	53 232	67 539	70 520	70 436	82 973	86 091

Source: Levy 2006, p.26.

The total annual budget of the program in 1999 was around US$777million, equivalent to just under 20 percent of the federal poverty alleviation budget or 0.2 percent of gross domestic product (GDP) (Skoufias 2004, p.1). In 2012 Progresa manages the highest budget granted to a program by the federal government and was allocated approximately US$5.3 billion for the year (Oportunidades 2012c).

4. Progresa-Oportunidades: Evaluation

Much of the information about the design and evaluation of Progresa is drawn from technical notes published by the Instituto Nacional de Salud Pública and Coordinación Nacional de Programa de Desarrollo Humano Oportunidades, SEDESOL (2005), on the general evaluation design in rural areas, and a final report presented by researchers from the International Food Policy Research Institute (IFPRI 1999). The latter evaluated the selection of households in rural areas for the program. Since Progresa has been subject to many evaluations, only the evaluation designed to measure its impact for the first three years, 1997-2000, will be reviewed.

4.1 Experimental Evaluation Design

Progresa has been a popular subject for evaluation over the years. Key indicators are used to measure the effect of the program on education, health and nutrition conditions, all causes and symptoms of poverty. The program's first phase was implemented as a randomized social experiment for evaluation purposes. The evaluation under discussion relates to this phase.

As discussed in section 2, social experiments randomly assign participants to either a treatment or control group. This ensures that each participant has the same opportunity to be assigned to either group. Due to the nature of random allocation it is correct to assume that there will be no difference in characteristics between the two groups and if there is, it is due to chance not bias. In the case of Progresa, randomization occurred first and foremost at a geographic level, targeting an initial group of participating localities. The localities were randomly allocated into either a treatment or control group, and all households living in treatment localities were identified as being eligible or not eligible to receive program benefits based on their level of poverty. Once a locality had been randomized to a treatment group, eligible households were notified of their acceptance into the program. This avoided a situation whereby households, living within the same community, were separated into treatment and control groups. The randomized design was validated by comparing characteristics of participants in the treatment and control groups, before receiving services (Schultz 2004). Notes on the methodology used for the rural sample by SEDESOL (2005 pp.10-11) state that the initial evaluation sample consisted of a total of 506 localities. Of these 320 localities made up the treatment group, while the remaining 186 became control localities. Schultz (2004 p.202), conveys that the 186 control localities became part of the program in the fall of 2000.

There are three stages in the selection of families as beneficiaries. Skoufias, Davis and Behrman (1999) describe them as follows:

1) First, a geographic targeting is carried out in which the localities with the greatest deprivation are selected, taking into account access to basic education and health services.
2) Second, in all of the selected localities, a survey of socio-economic information is carried out which is used to identify the beneficiary families.
3) Finally the list of selected families is presented to the community. An assembly is arranged in the community where the list of selected families is made public and feedback is collected about which families have been included or excluded incorrectly. Due to the minute numbers of households disputed, this stage won't be discussed any further.

Geographic Targeting of Localities

A Deprivation Index was used to rank localities into deprivation categories. Population and housing data for the localities was obtained from the 1990 XI General Population and Household Census, and the 1996 Population and Household Count and the information was used to compare the localities to the index. Localities were selected according to their level of deprivation thus localities with high or very high deprivation levels (those with a high proportion of households in extreme poverty) were given priority for inclusion in the program. Targeted localities were also required to have access to educational and health facilities. Where services were not located in the same community, availability and quality of roads were taken into account to consider the ground access to other localities with the services. An additional criterion was used for selecting localities in the first stages of the program whereby localities had to have more than 50 but less than 2,500 persons (SEDESOL 2005, p.5).

Identification of Eligible Families within Localities

Once highly deprived localities had been selected, the second stage of targeting followed which centred on evaluating the poverty conditions of the households within these localities. Progresa was concerned with targeting families living in extreme poverty. Accordingly, the methodology for identifying eligible families consisted of a questionnaire designed to capture the social and economic conditions of households. This information is then evaluated against a points system which defines the household's poverty condition. Selection for eligibility to receive the program's benefits followed.

Progresa establishes contact with the person (woman) who will be responsible for receiving benefits. At the same time that they are informed of being chosen as beneficiaries, the program and how it operates is explained and official identification and registration forms for schools and clinics are distributed. Each beneficiary must sign or put her fingerprint on a copy of the original documents. To keep receiving the benefits beneficiaries are informed of their responsibility to adhere to the program's requirements. Lastly they are made aware that they cannot receive benefits from any other federal program that provides equivalent benefits in the areas of education, health or nutrition.

4.1.1 Advantages of Randomization

Selection Bias

Selection bias occurs when individuals who participate in a program are different from individuals who do not participate in terms of unobservable characteristics. Progresa addressed selection bias in their evaluation through randomization. As described, this was done on the locality level for rural areas. Once the 506 localities had been selected and after meeting the inclusion criteria, 320 were randomly allocated to a treatment group, and the remaining 186 were considered the control group. Within the treatment group of localities, a poverty criterion was applied to determine which households would be eligible for benefits. The purpose of Progresa's method for identifying families' eligible to the program was to identify the set of socio-economic characteristics of the household and its members which determined whether or not the household lived in extreme poverty. A household was defined as the set of persons who lived together inside the dwelling (whether or not they are related), shared household expenditures and prepared food in the same kitchen (Skoufias, Davis and Behrman 1999, p. 35). Information on the socio-economic status of each household was collected through a questionnaire known as Encuesta de Características Socioeconómicas de los Hogares (ENCASEH), designed to obtain data on household characteristics as well as the well-being of its members. It took into account multiple factors which intervene in determining whether a household is poor or not. The following table summarizes the main topics included in the survey.

Table1.6: *Main Topics included in the ENCASEH Survey*

TOPICS OF THE SURVEY
1. Demographic characteristics of the household
2. Human capital of the household members
3. Economic activities of the household members
4. Income of the household members
5. Condition of the dwelling
6. Availability of services
7. Ownership of goods

Source: SEDESOL 2005, p.12.

The variables used in the questionnaire were selected after a detailed analysis of socioeconomic data sources at the national level, and selecting the socio-economic characteristics to evaluate the poverty condition was done through an extensive process in which the appropriateness of the link to poverty of each variable was analyzed until the best set of variables, which identified the poverty condition, was kept. This set of characteristics did not change significantly within regions and so poverty comparisons were consistent in that two individuals or households with the same level of consumption were treated the same, irrespective of the region they lived in. A common resident of the household could provide information for the questionnaire as long as they were 15 years of age, and could provide information on other household members. Ultimately the selection of beneficiary families was based on the analysis of the information acquired from each household in the selected locality.

The analysis of information involved a standardized process, at the national level, which evaluated the socio-economic data using a system of points. The system allows for the identification of beneficiary households based on unbiased criteria that are consistent to ensure the equal treatment of those in extreme poverty independent of where they live. The points reflect the instability of the economic resources of the family according to a set of basic indicators. The points take values between 0 and 100 with the lowest value associated with a higher propensity to be poor. The method of selecting beneficiary households begins with an initial approximation of the poverty condition through the use of the poverty line.[7] A first approximation of the families who live in extreme poverty is done using family income and the poverty line, which classifies households into two groups. After this a second classification is carried out using the set of social and economic indicators obtained from ENCASEH. From this multidimensional approach, it is possible to classify a household either in extreme poverty or not in extreme poverty. Thus all participants had the same known probability of being allocated to either group and it can be concluded that those who were allocated benefits did not differ from those who were not allocated benefits.

[7] See Appendix 2 for an explanation on how the poverty line is obtained.

Furthermore the evaluation design of Progresa satisfactorily addressed selection bias because no alternatives to the treatment were available. Progresa was implemented to replace all other poverty alleviating programs which meant that no other options for the same kind of services existed in the treatment locality. Households were not permitted to receive equivalent benefits from any other federal program if they were beneficiaries of Progresa.

Communication of Results

Burtless (1995) reminds us that an advantage of social experiments is that results can be communicated in straightforward language. Individuals without any formal training in economics or statistics are able to access the results which increase the chances of broad political acceptance. Progresa survived the crucial transition from President Zedillo to President Fox in December 2000, based on overwhelmingly supportive results from the first set of evaluations of the program. The results were presented to the incoming administration by officials from the outgoing administration after they had been made public a few months earlier. Evaluation generally provides evidence of accomplishment of goals and in this case, contributed to the continuity of the program due to the backing of the new President. Support from the president translated into continued funding for the program. Moreover in 2001, 750,000 additional households were added to the program and educational grants were extended for three more years to cover secondary school (Levy 2006, p.113). Some of the additional households included those that had been in the control group during the initial phase of the experiment. In these first years of the program a decision was made to have the program keep a low profile with regard to broader public opinion (Levy 2006, p.110). For the evaluation period considered in this paper, geographic mobility amongst localities did not play a role and therefore was not a concern for the experiment. Levy speaks about the positive opinions around these results embraced by international financial institutions like the Inter-American Development Bank and the World Bank. This made the information independent from any political affiliation and thus credible to the Fox administration. The straightforward way in which these results were able to be presented to a variety of audiences meant they were acceptable results and not subject to a variety of qualifications about their validity. The focus could then be on public policy and not scientific debate.

4.1.2 Threats to Experimental Design

Errors in the process of gathering the sample can occur inadvertently or on purpose. For the evaluation design of Progresa, Skoufias et al. (1999, p.39), found the method of selecting beneficiary families to be transparent while at the same time rigorous and objective. The procedure for selecting these families used a classification rule which originated in Progresa's design and was derived from the information of the households to be classified. Additionally the method permitted new households, who did not report their income level, to be classified according to their other socio-economic characteristics. This had the effect of increasing the reliability of the selection process by reducing the risk of errors in the classification of households due to incorrect or inaccurate income measurement. The procedure is applied using consistent criteria without instituting a predetermined number of families who "should" be Progresa beneficiaries (Skoufias et al. 1999, p.39). In contrast these numbers are derived from the procedure of identification of households in extreme poverty.

Procedures to ensure quality of data being collected were constantly supervised. The process of data collection (with reference to socio-economic data collection from households) was supervised through control of sample devices of coverage and quality control, carried out by groups independent from those in charge of carrying out the questionnaire. Additionally the municipal authorities and the local representative provided support by helping to identify the geographic borders of the locality and authorizing the activities of the groups in charge of carrying out the questionnaires (Skoufias et al., 1999).

As Heckman and Smith point out, there can be challenges in the practical running of a random assignment evaluation. In Progresa's case those in central management were knowledgeable and understanding of the program and responsible for its administration. Finance ministry officials ran the program.

4.1.3 Disadvantages and Limitations

Institutional factors

Before Progresa could be implemented, localities were required to have access to educational and health facilities. If these were not situated within the locality they had to be within a reasonable distance. Notably almost 70% of all selected localities did have access to health and education services (SEDESOL, 2005). However some of Mexico's poorest residents lived in remote areas which did not have these facilities and thus were not recipients of the program.

Progresa enjoyed the advantage of having support from the highest management in the country, including the president and officials from the ministry of finance. The government was an advantage for Progresa and in its implementation.

Partial Equilibrium Results

Social experiments have been criticised for not reflecting the general equilibrium effect of a service treatment. In Progresa's case, as the implementation stage was on a relatively small scale, the partial equilibrium results were near the same as results as the general equilibrium effects. One could suppose that as the program was scaled up, general equilibrium results could include a decrease in the labour supply as more children are enrolled in school. This could place upward pressure on wages meaning instead of spending money on labourers this money could be used to invest in the capital of the business (farming). Once children have finished their schooling they are more likely to leave their communities in search of better opportunities, placing downward pressure on wages as labour supply increases. A spillover effect of children receiving better nutrition and more regular health checks may imply less transmission of diseases than evident from partial equilibrium results. A last consideration would be the change in family dynamics as a result of mothers receiving the benefits. Perhaps they would feel empowered and sure of their position in the household. This could bring about a shift in the balance of power in the relationship between husbands and wives with wives having more of a say in how the families' resources are spent.

Cost

Evaluation was always considered essential to the program because of its value in exposing what was faulty and needed to be corrected or adjusted, so that resources were increasingly used in effective and efficient ways to achieve the program's goals.

At the start of implementation, the cost of collecting data and building a database of household and community surveys was taken from the available budget. This did decrease the amount available for poverty alleviation however the return from investing in an evaluation early on, allowed for changes to be made to streamline the use of resources; resulting in what is now the principal poverty alleviating strategy of the Mexican government and is a model for similar programs elsewhere in the world.

Ethics

In a situation of limited resources the most pressing question is how to use them most efficiently and effectively. As it was not possible for treatment to be available to all eligible applicants, a fair determinant of which families were to receive benefits first was through randomization. No harm was done to the control group as treatment was otherwise not available. No tax hike was imposed to generate additional revenue for Progresa because it was designed to replace all food subsidy programs. Resources were made available by the federal government from reorganising the budget for social policies, and shifting resources from general food subsidies to targeted subsidies (Levy 2006 pp.83-87). Furthermore the control group began receiving benefits after the first three years.

For the ENCASEH survey, with regard to confidentiality, each household and its members were assigned an identification number to allow for confidentiality of the information collected and this information was saved in magnetic files for processing.

Evaluation of socio-economic data is documented and saved for all the stages of identifying of beneficiary families, a process which is transparent and does not depend in any way on any individual's political affiliation. It is worth remembering Progresa was taking place in a context of rapid democratization where transparency and accountability were highly valued principles. Evaluations of Progresa were also made available to the public.

Communities within localities were given a voice in the program and consulted on the different aspects of the program thereby establishing relationships between Progresa officials and communities. Once families had been selected for the program the community was given the opportunity to give feedback to the program administrators, particularly if a family should or shouldn't have been included.

5. Progresa-Oportunidades: Results

5.1 Validity of Randomization

Before looking at the effects of the program, researchers set out to determine whether the control and treatment groups were in fact randomly assigned. The purpose of this was to detect any deviations early on and take them into account during the course of the evaluation (Behrman and Todd 1999). Since random assignment implies that different variables belonging to treatment and control groups are, on average, equal before the program is administered, then it can be evaluated by comparing the characteristics of treatment and control group households, again prior to having received any program benefits. Any observed differences between the groups can undergo testing to determine if they are statistically significant.

Behrman and Todd (1999) compared the treatment and control groups in terms of geographic location patterns, age distributions, educational attainment distributions, access to health care distributions and income distributions. On top of this they used formal tests of equality between the groups amongst a distribution of various characteristics. This established whether to accept or reject the hypothesis that both groups were equal in characteristics.

The following table compares the age distribution of children below 17 years old in both the treatment and control groups. The patterns appear similar and show no evidence of any systematic deviation within the two groups.

Table 1.7: *Age Distribution of Children < 17 years old by Treatment/Control Status*

(Based on ENCASEH97)

AGE IN YEARS	TREATMENT	CONTROL	TOTAL
<1 Year Old	5.22	5.23	5.22
1	5.71	5.42	5.60
2	6.25	6.23	6.24
3	6.48	6.54	6.50
4	6.91	6.80	6.87
5	6.48	6.66	6.55
6	6.74	6.57	6.67
7	6.46	6.73	6.56
8	6.65	6.44	6.57
9	6.04	5.67	5.90
10	6.28	6.48	6.36
11	5.89	6.16	5.99
12	5.88	5.88	5.88
13	5.24	5.45	5.32
14	5.22	5.21	5.22
15	4.79	4.81	4.80
16	3.77	3.73	3.76
Total	100.00	100.00	100.00

Adapted from Behrman and Todd 1999, p.17.

Another table from Behrman and Todd's investigation (1999) summarizes the educational attainment distributions for treatments and control groups.

__Table 1.8:__ Educational Attainment of Household Heads by Treatment/Control Status

(Based on ENCASEH97)

LEVEL	TREATMENT	CONTROL	TOTAL
Incomplete primary	0.30	0.64	0.34
Complete primary	67.35	69.13	68.01
Incomplete secondary	24.72	22.96	24.06
Complete secondary	1.78	2.10	1.90
Incomplete normal basic, preparatory or professional	4.81	4.26	4.60
Complete normal basic, preparatory or professional	0.57	0.41	0.51
Post-graduate	0.47	0.51	0.49
Total	100.00	100.00	100.00

Adapted from Behrman and Todd 1999, p.18.

Although most of the distributions appear to be similar (completed primary or incomplete secondary) a formal test of equality rejected the hypothesis that both groups were equal in terms of characteristics. The authors found that when tests were performed on household level data, more rejections occurred than when they were performed on locality level data. Their conclusions were that these rejections were due to large sample sizes at the household level where equality, even on the basis of minor differences, was rejected. Generally from the evidence presented from their analysis, Behrman and Todd found the randomization process to be adequately done based on similarities in characteristics between the groups in terms of age, education, access to health care and income. There were no systematic differences. Formal tests of equality between the distributions of various characteristics generally did not reject the hypothesis of equality when the test was performed on locality means (Behrman and Todd 1999, p.8).

Further evidence that randomization was successfully carried out in allocating localities to treatment and control groups can be found in Paul Gertler's report to the IFPRI (2000).[8] Before evaluating health outcomes he checked the success of randomization by using a baseline sample of children aged 0-5 years containing a variety of variables, to compare between the treatment and control groups. The same sample was used for the majority of analyses done by researchers.

Table 1.9: *Descriptive Statistics at Baseline for Children Age 0-5*

	Treatment		Control	
	Mean	St. Error	Mean	St. Error
Ill last month (=1)	0.306	(0.461)	0.298	(0.458)
Nutritional Monitoring Visits Last Month	0.219	(0.217)	0.219	(0.206)
Age	2.753	(1.667)	2.746	(1.701)
Male (=1)	0.394	(0.489)	0.376	(0.484)
Father's years of Schooling	1.111	(1.093)	1.050	(1.086)
Mother's Years of Schooling	1.047	(1.056)	1.016	(1.079)
Number of Siblings	1.996	(1.881)	1.783	(1.831)
Eldest Child (=1)	0.916	(0.277)	0.930	(0.256)
Labor & Non-labor Income	4.939	(0.896)	5.094	(0.814)
Non-labor Income	2.883	(1.152)	2.898	(1.157)
Male Agricultural Wage	23.658	(7.573)	23.556	(6.935)
Female Agricultural Wage	20.984	(7.184)	21.044	(6.807)

Adapted from Gertler 2000, p.23.

As the table shows, Gertler found no difference in illness rates or number of visits to clinics for nutrition monitoring, at baseline, between treatment and control groups. Given that standard error measures the accuracy of the mean, little difference between family demographics or economic status was found, and there seemed to be no difference in labour markets as the

[8] Gertler worked as a consultant for the IFPRI conducting research on the impact of Progresa on an individual's health.

agricultural wages were the same across treatment and control localities. He concluded that because the analysis had shown randomization to be successfully carried out between the groups on observed characteristics, it was likely that the groups were balanced on unobserved characteristics as well (Gertler 2000, p.7), even though there is no way of really knowing this.

Knowing that the randomization process was sound, researchers could confidently infer whether changes in the observations of individual outcomes were due to the program or other factors. We now examine some findings in health outcomes.

5.2 Findings

The IFPRI was asked to assist Progresa to evaluate the effects of the program, in order to determine whether it was meeting its objectives. The evaluations were based on the pre-intervention baseline study, ENCASEH, and four follow-up surveys done at six month intervals of the same households conducted over a two year period (Gertler 2000). The surveys were supplemented by focus groups and workshops with beneficiaries, local leaders, Progresa officials, health clinic workers and teachers. The method used to evaluate the information from the surveys involved comparing mean values on given variables between the treatment and control groups. In order to be as accurate as possible regression methods were employed to control for observable characteristics that might also have an effect on that variable. The data made it possible for evaluators to measure the impact of the program on a large number of variables of interest (Levy 2006, p.38). Dr. Emmnauel Skoufias, the coordinator of the Progresa – IFPRI evaluation, summed up the results of Progresa in his foreword (Skoufias 2000).

> *"In brief, the findings of IFPRI's evaluation are that after just three years, the poor children of Mexico in the rural areas where PROGRESA is currently operating are more likely to enroll in school, are eating more diversified diets, getting more frequent health care and learning that the future may look quite different from the past."*

Health Outcomes

Most of the analysis in this section draws on Gertler's final report (2000) on the impact of Progresa on the health outcomes of individuals. The analysis will focus on three hypotheses he proposed to measure Progresa's impact on health outcomes.

The first hypothesis was that there was an increase in visits to public clinics by beneficiaries (Gertler 2000, p.8). One would assume that there would be an increase because of the requirement of Progresa to regularly attend clinics and the cash transfers given for nutrition could be used to purchase medical care. However if Progresa's interventions have succeeded in preventing illness, there should be a decrease in visits to clinics because of a reduction in illness. Alternatively a decrease in visits could indicate that the number of visits by beneficiaries already outnumber those required by Progresa. Gertler brings up a further issue concerning an increase in visits to public clinics by asking whether this increase comes from individuals substituting private provider visits for public clinic visits. If this is the case the impact of the program on health outcomes would be less because there would be no new utilization of public clinics; public clinic visits would merely have been substituted for from private provider care. Health effects would be determined by the difference in quality between public and private sectors rather than the difference between the quality of public treatment and self-treatment (Gertler 2000, p.8). To find out whether an increase in visits to clinics was new utilization or a substitution from private to public care, Gertler examined the impact of Progresa on total utilisation of all provider types and then disaggregated the impact by provider type: public clinic; public hospital; private provider. He needed to show that the total increase in utilization of clinics was the same as the increase in public clinics and that there was no decrease in visits to private providers, to be able to accept the hypothesis.

As beneficiaries have to visit public clinics, Gertler's first test for the hypothesis was one of compliance, that of beneficiaries meeting the program's requirements. Gertler used data from the administrative records of public clinics[9] and records kept by Progresa on the number of beneficiary families incorporated into the program every month and in each clinic, for his investigation. Approximately two-thirds of the clinics were in treatment localities while the rest were in the control locality. The average daily visits to a public health clinic were measured for both treatment and control localities by month over time. Gertler found the visit rates to be almost identical until the latter part of 1997 when Progresa began being introduced to treatment localities. At this time visit rates to clinics in treatment localities were on average higher than in control localities and the difference grew over time as more treatment localities were provided benefits (Gertler 2000, p.9). Table 2 below corresponds by measuring average daily visit rates for treatment and control localities by year.

Table 2: *Total Daily Consultations per Clinic Means (with Standard Errors)*

	Treatment	Control	Overall
1996	9.11	9.13	9.12
	(7.98)	(8.25)	(8.06)
1997	10.75	10.35	10.63
	(9.10)	(9.14)	(9.12)
1998	12.84	11.48	12.41
	(11.32)	(9.80)	(10.88)
Overall	10.0	10.32	10.72
	(9.69)	(9.69)	(9.52)

Adapted from Gertler 2000, p. 23.

In 1996, the year before Progresa was implemented; average visits to clinics were the same in treatment and control localities. These change in 1998 when Progresa becomes

[9] There were 3 541 clinics and the data included monthly information from January 1996 to December 1998 (Gertler, 2000, p.8).

operational in all treatment localities. Visit rates were found to be 12% higher in treatment localities than in control localities. In addition to this Gertler used a difference-in-difference estimator to compare the change (before and after intervention) in visits per day in treatment localities with the change in control localities. Simultaneously he controlled for characteristics that do and do not change over time and are common to both groups, like changes in the size of the clinic's service area and the number of beneficiary families in the service area. The results show a difference in the mean number of daily visits for treatment and control localities. They indicated an increase of about 2.09 more visits per day to clinics in treatment localities than in control localities, in other words double the amount than families not receiving the program (Gertler 2000, p.10). Both these tests are consistent with Gertler's hypothesis that there was an increase in visits to public clinics by beneficiaries. This increase, though small was positive, and represents new visits for preventative care by beneficiaries who had not previously visited health care clinics (Levy 2006, p.49).

To investigate the total utilisation of all provider types, Gertler used data from the household surveys which contained information on individuals' utilization of health care. This was measured by questions pertaining to number of visits to a public hospital, a health clinic, a private hospital, a private doctor, a med-wife, herbalist or traditional doctor, and a pharmacy (Gertler 2000, p.10). There was no baseline information. Generally he found health care utilization by the rural poor to be extremely low. Disaggregating into geographic areas he found a higher utilization rate for people in treatment localities than in control ones. In all age groups he found the majority of people utilized public clinics.

Table 2.1: *Means and Standard Errors of Visits to Medical Care Providers*

Age	Sample	Mean Visits per Month								Sample Size
		Total		Public Clinics		Hospitals (inpatients)		Private Providers		
<3	Treatment	.081	(1.42)	.066	(1.41)	.012	(0.09)	.003	(0.13)	5 110
	Control	.115	(1.62)	.079	(1.58)	.011	(0.03)	.025	(0.27)	4 110
3 to 5	Treatment	.097	(1.29)	.075	(1.27)	.005	(0.10)	.017	(0.18)	6 443
	Control	.068	(0.41)	.046	(0.33)	.004	(.013)	.018	(0.21)	5 717
6 to 17	Treatment	.041	(1.05)	.034	(1.05)	.002	(0.07)	.005	(0.10)	28 526
	Control	.027	(0.68)	.017	(0.64)	.001	(0.05)	.008	(0.21)	25 259
18 to 50	Treatment	.071	(0.95)	.050	(0.91)	.005	(0.11)	.016	(0.26)	26 702
	Control	.071	(0.95)	.050	(0.91)	.005	(0.11)	.016	(0.26)	26 702
>51	Treatment	.139	(1.47)	.095	(1.33)	.006	(0.11)	.038	(0.49)	6 927
	Control	.139	(1.47)	.095	(1.33)	.006	(0.11)	.038	(0.49)	6 927

Adapted from Gertler 2000, p.25

Noteworthy measures from the table show that for 3-5 year olds and 6-17 years old, the total increase in total visits is equal to the increase at public clinics and indicates no impact on private provider visits. This represents new utilization from Progresa's requirement for checkups. For 18-50 year olds and for those over 50, the impact of Progresa on total visits is 50% and 20% less than the impact on visits to public clinics. Again there was no impact on visits to private providers. This suggests that the increase in utilization at public clinics was not from substitution out of the private sector for these groups, but rather new utilization for preventive purposes (Gertler 2000, p.12).

The second hypothesis Gertler proposed was that Progresa's preventative activities reduced illness therefore reducing the demand for curative care (Gertler 2000, p.10). If this held true then total utilization (curative and preventative) of health facilities would fall. The focus of the examination was to see whether the total visits and hospital inpatient visits were lower for beneficiaries. The method used tests based on the following conditions: first to test if total visits were lower because a reduction in curative visits due to people being healthier is greater than an increase in preventative services. Second, hospitals do not provide preventive services and so a reduction in visits to hospitals would suggest that Progresa reduced serious illness. He also analysed nutrition monitoring visits to clinics and found an increase of 30-60% for children aged 0-2 and an increase of 25-45% for children aged 3-5, in treatment localities (Gertler 2000, p.13).

Gertler presents the results from his tests in his final report. To summarize he found that for 0-2 year olds the effects suggest that total visits fell by 37% even though the estimate is not significantly different from zero. However he found a significant 58% reduction in hospital visits for this age group pointing toward a significant reduction in major illness. Similarly he found a large reduction in hospitalization for the over 50 group. This suggests that Progresa had a positive impact on health status which is consistent with the hypothesis that Progresa's preventative activities reduced illness therefore reducing the demand for curative care.

The third and final hypothesis Gertler proposed was that Progresa lowered the incidence of severe illness in both children and adults. He examined both the health status of children and of adolescents and adults in his analysis. To examine outcomes on child health he used difference-in-difference estimates controlling for household per capita income. The reason for controlling for an observed characteristic was because of Behrman and Todd's (1999) findings that detected some differences when means were compared at the household level. The data used was information collected from the same group of children for the baseline and follow-up surveys.

__Table 2.2:__ Child Illness Rates by Age and Treatment and Control Status

		Treatment	Control
Age 0-2	Baseline	0.402	0.406
	8 months Post Baseline	0.284	0.366
	15 months Post Baseline	0.193	0.241
	20 months Post Baseline	0.194	0.246
	Sample Size	5445	2171
Age 3-5	Baseline	0.280	0.263
	8 months Post Baseline	0.206	0.270
	15 months Post Baseline	0.127	0.161
	20 months Post Baseline	0.097	0.127
	Sample Size	11370	4066

Adapted from Gertler 2000, p.28.

In the baseline survey, illness rates were the same across treatment and control localities. Illness rates in both groups fall over time although the illness rate for the treatment group falls faster than the control group. Regression methods measured the impact of Progresa on the probability of a child getting ill and the results were negative and statistically different from zero. Progresa lowered illness rates for beneficiaries' age 0-2 years old by about 12% lower than baseline illness. The illness rate for beneficiaries' age 3-5 years old decreased by 11% lower than baseline illness (Gertler 2000, p.14). It was also found that adding income as a variable does not change the result, suggesting that the impact on children's health is not a result from cash transfers.

Even though Progresa was not designed to specifically improve adult health, there was reason to believe that it did. All household member receiving the program were required to have an annual health care visit and Hoddinott and Skoufias (2000) found that a large part of the cash transfers were used to increase food consumption in terms of quantity and quality. Both these factors are related to the health status of an individual. To capture information about the health of adolescents and adults, questions were asked in the last two rounds of the survey regarding their health status. The variables are presented in the table below, along with means and standard deviations.

Table 2.3: *Means and Standard Errors of Health Status Measures*

	Age 6-17		**Age 18-50**		**Age 51+**	
	Treatment	Control	Treatment	Control	Treatment	Control
Days of difficulty with daily activities due to illness in last 4 weeks	0.084	0.087	0.287	0.347	1.875	2.271
	(1.266)	(1.164)	(2.496)	(2.756)	(6.648)	(7.274)
Days incapacitated due to illness in last 4 weeks	0.081	0.071	0.248	0.288	1.601	1.961
	(1.321)	(1.107)	(2.348)	(2.506)	(6.106)	(6.783)
Days in bed due to illness in last 4 weeks	0.043	0.045	0.172	0.185	1.089	1.355
	(0.901)	(0.854)	(1.894)	(1.970)	(5.124)	(5.630)
Kilometres can walk without getting tired			5.497	5.085	3.273	3.018
			(4.056)	(3.474)	(3.180)	(3.361)
Sample size	28 526	25 259	26 702	26 388	6 927	8 472

Adapted from Gertler 2000, p.29.

The results found no effect of the program on individuals aged 6-17, Gertler comments that this is not surprising since its generally a healthy group to start with (Gertler 2000, p.15).

However, for the age group 18-50 a significant reduction in the number of days of difficulty with daily activities due to illness is found as well as a significant increase in the number of kilometers able to walk without getting tired. Specifically Progresa beneficiaries have 19% fewer days of difficulty due to illness than non-beneficiaries and are able to walk about 7.5% without getting tired. For those over 50, beneficiaries have significantly fewer days of difficulty with daily activities, days incapacitated, and days in bed due to illness than non-beneficiaries. In addition they are able to walk more kilometers without getting tired. Again specifically, beneficiaries had 19% fewer days of difficulty with daily activities, 17% fewer days incapacitated, 22% fewer days in bed, and were able to walk about 7% more than non-beneficiaries (Gertler 2000, p.15).

Altogether these results are consistent with the hypothesis that Progresa lowered the incidence of severe illness in both children and adults.

6. Conclusion

The major advantages of social experiments are that they address selection bias through random assignment, and the results are easily communicated. Skoufias (2000) maintains that random assignment lends exceptional strength to the evaluation process. It ensures every person of a target population has the same, nonzero chance of being selected for the treatment group. As a result individuals gathered for the experiment will be of similar motivations. In addition no alternative for the same kind of services should be available. Random assignment supports the internal validity of the experiment and the results are as close as we can get to estimating causal impact. The relative simplicity of social experiments means that results can be communicated in straightforward language. Not only are the results acceptable to social scientists but are also accessible to individuals with no formal training in economics or statistics. These may include policymakers. The benefit of this is an increased chance of broader political acceptance of the findings which implies that results are used for changing public policy.

Heckman and Smith (1995) argue that the major limitations of social experiments relate to the divergence between theoretical capabilities of experimentation and the practical results. When the running of the experiment does not match theory, it risks internal threats to its design. Furthermore Burtless and Orr (1986) reason that the fundamental ethical principle behind experimentation should be that the expected net benefits to participants of both the treatment and control groups not be less than those they could expect in the absence of the experiment. This condition is met in an experiment which provides a beneficial treatment that would otherwise not have been available. The treatment group is better off because of their participation, and the control group is no worse off because of their participation. Also there are circumstances where it is appropriate to deny services to a control group. Every experiment takes place within a budget, and therefore it may not be possible for the treatment service to be available to all eligible applicants. Random assignment would then be a fair way to select the individuals.

Progresa capitalized on the advantages in this way. Its evaluation was conceived from the beginning to be part of the design of the program. Random assignment was carried out at the local level, and therefore every person in each locality had same, nonzero chance of being selected for the treatment group. It solved the problem of not being able to simultaneously observe households in the alternate state of no treatment (Gertler 2000, p.6). The validity of the procedure of random assignment was subject to scrutiny and found to be adequately achieved. Thus outcomes supported the internal validity of the experiment and researchers could carry out evaluations with confidence in the randomization process. Progresa was implemented to replace all other poverty alleviating programs so no similar programs with the same kind of services existed. The positive results from the evaluations of Progresa between 1997 and 2000 were impressive. They were made public and communicated to the incoming administration by officials from the outgoing administration. Evidence of the accomplishment of goals had been given and resulted in the continuation and expansion of Progresa by the new government.

Progresa minimized the limitations of social experimentation by making sure the theoretical capabilities translated into the practical running of the program. This was made possible because central management had knowledge and understanding of the program and were responsible for its administration. Moreover Progresa had the support of the highest management in the country. In terms of ethics, Progresa was constrained by a budget and so it was not possible to make treatment available to all eligible participants. In this case random assignment was a fair way to select beneficiaries. Progresa was providing beneficial services which would not otherwise have been available thus met the condition that the expected net benefits to participants of both the treatment and control groups, not be less than those they could expect in the absence of the experiment. Although the treatment group was better off for their participation the control group was no worse off for theirs. No harm had been done.

This paper has examined how social experimentation can show in what way policies work. The Progresa-Oportunidades program was used as a case study because it engaged social experimentation as an evaluation tool, and was able to marry the theoretical aspects of experimentation with the practical running of one. The importance of the evaluation was that it played a major role in ensuring the program not only survived a change in government, but was expanded in 2000. As such Progresa is now the principal anti-poverty strategy of the Mexican government and at the international level, has been a model for similar programs in a number of countries in Latin America (Parker and Teruel 2005).

7. References

Becker, G.S. (1999, November 22). 'Bribe' Third world parents to keep their kids in school. *Business Week (Industrial/Technology Edition)*. Retrieved from http://home.uchicago.edu/gbecker/Businessweek/BW/1999/11_22_1999.pdf

Behrman, J.R. (2007). *Impact Assessment Discussion Paper No. 27: Policy-Orientated Research Impact Assessment (POIRA) Case Study on the International Food Policy Research Institute (IFPRI) and the Mexican Progresa Anti-Poverty and Human Resource Investment Conditional Cash Transfer Program.* Washington, D.C.: International Food Policy Research Institute.

Behrman, J.R and P.E Todd. (1999). *Randomness in the experimental samples of PROGRESA (education, health, and nutrition program)*. Report submitted to Progresa. Washington, D.C.: International Food Policy Research. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Burtless, G. (1995). The Case for Randomised Field Trials in Economic and Policy Research. *Journal of Economic Perspectives*, 9:2, 63-84.

Burtless, G. and L.L. Orr. (1986). Are Classical Experiments Needed for Manpower Policy? *The Journal of Human Resources*, 21:4, 606-639.

Coady, D.P. (2000). *Final Report: The application of social cost-benefit analysis to the evaluation of Progresa.* Report submitted to PROGRESA. Washington, D.C.: International Food Policy Research. Retrieved November 5, 2012, from http://evaluacion.oportunidades.gob.mx:8010/441c7c1a3d30adf64e0e724174a9d527/impacto/2000/ifpri_2000_coady_scba.pdf

Davis, B. (2003). *Choosing a Method for Poverty Mapping*. Rome: Food and Agricultural Organisation of the United Nations. Retrieved September 24, 2012, from ftp://ftp.fao.org/docrep/fao/005/y4597e/y4597e00.pdf

Fukuyama, F. (2007, February 1). Commentary: Keeping up with the Chavezes. *The Wall Street Journal*: p. A17.

Gertler, P. (2000). *Final Report: The impact of Progresa on health*. Report submitted to Progresa. Washington, D.C.: International Food Policy Research. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47:1, 153-161.

Heckman, J. and J. Smith. (1995). Assessing the Case for Social Experiments. *Journal of Economic Perspectives*, 9:2, 85-110.

Hoddinott, J., and E. Skoufias. (2000). *Final Report: The impact of PROGRESA on consumption.* Report submitted to Progresa. Washington, D.C.: International Food Policy Research. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Instituto Nacional de Salud Pública Coordinación Nacional de Programa de Desarrollo Humano Oportunidades, SEDESOL, (2005). General Rural Methodology Note. Retrieved October 17, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/notes_rural.php

Krueger, A. B. (2002, May 9). Economic Scene: A model for evaluation the use of development dollars, south of the border. *New York Times*. Retrieved from http://www.nytimes.com/2002/05/02/business/economic-scene-model-for-evaluating-use-development-dollars-south-border.html?ref=alanbkrueger

Levy, S. (2006). *Progress against Poverty: Sustaining Mexico's Progresa*-*Oportunidades Program.* Washington, D.C.: Brookings Institution Press.

Oportunidades 2012a, "The Context of Poverty in Mexico" http://www.oportunidades.gob.mx/Portal/wb/Web/the_context_of_poverty_in_mexico, accessed October 13, 2012.

Oportunidades 2012b, "Design and Operation" http://www.oportunidades.gob.mx/Portal/wb/Web/design_and_operation, accessed October 17, 2012.

Oportunidades 2012c, "Budget" http://www.oportunidades.gob.mx/Portal/wb/Web/budget, accessed October 17, 2012.

Parker, S.W. and G.M. Teruel. (2005). Randomization and Social Program Evaluation: The Case of Progresa. *The Annals of the American Academy*, 599:1, 199-219.

Schultz, T.P. (2000). *Final Report: The impact of PROGRESA on school enrollments*. Report submitted to PROGRESA. Washington, D.C.: International Food Policy Research. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Schultz, T.P. (2004). School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. *Journal of Development Economics*, 74, 199-250.

Skoufias, E. (2005). *Progresa and Its Impacts on the Welfare of Rural Households in Mexico.* Washington, D.C.: International Food Policy Research Institute. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Skoufias, E., B. Davis, and J.R. Behrman. (1999). *Final Report: An evaluation of the selection of beneficiary households in the education, health, and nutrition program (PROGRESA) of Mexico.* Washington, D.C.: International Food Policy Research Institute. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Skoufias E. and B.E. Todd. (1999). *A report on the sample sizes used for the evaluation of the education, health, and nutrition program (PROGRESA) of Mexico*. Washington, D.C.: International Food Policy Research Institute. Retrieved October 11, 2012, from http://evaluacion.oportunidades.gob.mx:8010/en/docs/eval_docs_2000.php

Treasury Board of Canada Secretariat. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=15024&section=text#cha1, accessed September 1, 2012.

Weiss, C.H. (1998). *Evaluation: Methods for Studying Programs and Policies. Second Edition.* New Jersey, U.S.A: Prentice Hall.

World Bank 2012, "Evaluation Designs"

http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20188242~menuPK:415130~pagePK:148956~piPK:216618~theSitePK:384329,00.html

accessed November 18

Worthen, B., Sanders J.R. and Fitzpatrick L.J. (1997). Program Evaluation: Alternative Approaches and Practical Guidelines. New York, NY: Longman Publishers.

8. Appendices

8.1 Appendix 1

The Level and Composition of Transfers to the Beneficiary Families (million pesos, 2000 prices)

Transfer Components	1997	1998	1999	2000	2001 / Predicted
Education transfers	131.49	916.33	2,797.76	4,003.00	4,258.63
Scholarships	127.44	na	2,729.38	3,681.60	3,916.71
Materials	4.05	na	68.38	321.40	341.92
Food transfers	97.21	1,029.60	2,985.69	3,753.88	3,993.60
Health Component	112.87	1,858.44	1,305.83	1,461.70	1,555.05
Supplements	36.92	965.81	744.25	721.60	767.68
Other	75.95	892.63	561.58	740.10	787.37
Total Cash Transfers (Food+Education)	228.70	1,945.93	5,783.45	7,756.88	8,252.23
Food Share	0.5749	0.4709	0.4838	0.5161	0.5161
Scholarship share	0.4251	0.5291	0.5162	0.4839	0.4839
Total Transfers	341.57	3,804.37	7,089.28	9,218.58	9,807.28

Note: The numbers for 2001 are predicted using the food transfers for 2001 and applying the ratios from 2000.
Adapted from Coady 2000, p.87.

8.2 Appendix 2

The poverty line is obtained through the evaluation of monthly income per-capita as compared with the cost of a basic food basket. Monthly per capita income takes into account all members of the household aged 15 or more. The cost of the basic food basket corresponds to the Normative Food Basket, defined by the General Coordination of the National Plan for Depressed Areas and Marginalized groups and satisfies the minimum requirements to prevent malnutrition and diseases. The extreme poverty line is adjusted for inflation using the National Consumer Price Index, published by the Bank of Mexico.