



RE2-06-033

**Economic and
Sector Study Series**

**DEMAND-SIDE INCENTIVES FOR
BETTER HEALTH FOR THE POOR:
CONDITIONAL CASH TRANSFER
PROGRAMS IN LATIN AMERICA
AND THE CARIBBEAN**

**Amanda Glassman
Marie Gaarder
Jessica Todd**



December 2006

REGION II

Inter-American Development Bank

This document in discussion is not an official publication of the Inter-American Development Bank. The purpose of the Economic and Sector Study is to provide a mechanism for diffusion of selected analytical work undertaken by the department in support of its operational program at the country or sub-regional level. Opinions and judgments expressed in these studies do not necessarily reflect the view of Bank Management or member countries.

<http://www.iadb.org/publications>

PREFACE

Conditional Cash Transfer (CCT) programs are spreading rapidly throughout the developing world. Since 1997, seven countries in Latin America and the Caribbean (LAC) have implemented and evaluated CCT programs with health and nutrition components. These are the subjects of this paper and include Brazil's Bolsa Alimentacao/Bolsa Familia, Colombia's Familias en Acción, Ecuador's Bono de Desarrollo Humano, Honduras' PRAF, Jamaica's PATH, Mexico's Progres/Oportunidades, and Nicaragua's Red de Protección Social. Others are in the process of development in Argentina, Chile, Costa Rica, El Salvador, Panama and Paraguay, and around the world.

CCT programs provide cash to poor households, mostly in rural areas, upon household compliance with a set of health and education-related conditions. Expected immediate results include increased food consumption, school attendance and preventive health care utilization among the poor. Longer-term expected impacts are increases in the accumulation of human capital and associated returns in the labor market. These expectations arise from a set of implicit assumptions about how CCT programs effect change and influence behavior, and how induced behavior affects health status. Understanding the causal pathways that underpin an intervention is critical to evaluating how it works and what should be modified to improve effectiveness. This paper analyzes critically the program effect model with respect to health and nutrition outcomes, based on the LAC experience, and draws some conclusions regarding the use of CCT programs to improve health and nutrition status and to promote healthy behaviors.

Amanda Glassman, M.Sc., Health Financing Task Force/Brookings Institution; Marie Moland Gaarder, Ph.D., Social Development Specialist in the Social Programs Division II (RE2/SO2), Inter-American Development Bank (IADB); and Jessica Todd, Ph.D. candidate, American University, are the authors of this paper.

The authors thank Markus Goldstein, Rena Eichler, Ruth Levine and the members of the Center for Global Development Payment for Performance Working Group for their comments. The authors also thank Héctor Salazar-Sánchez, Division Chief (RE2/SO2) and Manuel R. Agosin, Consultant (RE2/RE2) for their guidance. Additionally, the research assistance of Lucia de la Sierra is also appreciated, and the support of Mariana Williman (RE2/SO2) and Miriam Pérez Fuentes (RE2/RE2) in the production of this document gratefully acknowledged.

Máximo Jeria
Manager
Regional Operations Department II
Central American Isthmus, Mexico, Haiti and Dominican Republic

Washington, D.C., December 2006

CONTENT

I.	Introduction.....	1
II.	Basic design features of CCT programs	2
III.	Understanding the demand and supply-side factors that determine utilization and effectiveness of health care services.....	9
IV.	Assumptions underlying the CCT program effect model and impact evaluation methodology.....	10
V.	Conclusions and recommendations	22
	A. OPORTUNIDADES (MX)	36
	B. RED DE PROTECCIÓN SOCIAL (NI)	37
	C. PRAF (HO).....	37
	D. FAMILIAS EN ACCIÓN (CO)	38
	E. PATH (JA).....	39
	F. Questionnaires	40

I. INTRODUCTION

Poor families in the developing world face resource constraints and other disincentives to the use of essential health care. Distance to health facilities, lost wages associated with illness, care-taking and care-seeking, facility fees and other out of pocket costs all contribute to limit access to health care and information by those who need it most, particularly preventive care. These costs also have an impact on the financial security of poor households, as out of pocket spending can drive poor families deeper into poverty.

Counteracting these disincentives requires a multi-pronged strategy aimed at both the supply and demand sides of a health system. This paper focuses on a type of demand-side transfer known as a Conditional Cash Transfer (CCT). CCT programs transfer cash to poor mothers to ease constraints on household investment in human capital development by reducing out-of-pocket expenditures and opportunity costs. CCT are also intended to induce health and nutrition behavior changes by conditioning the transfers on the use of preventive health services and attendance to health education talks. CCT payments can thus be considered demand-side payments for performance, where performance is understood as a healthy behavior.

The origins of CCT programs are diverse, some related to cushioning the effects of economic crises, others as a response to the perceived failures of existing safety nets. As a result, the central objective of the programs is to reduce poverty. A second set of objectives relates to human capital accumulation. Expected immediate results are increased food consumption, school attendance and preventive health care utilization among the poor. In the longer-term, CCT are expected to directly contribute to the accumulation of human capital and associated returns in the labor market through the reduction of malnutrition and the improvement of health and schooling completion rates, and thus, increased productivity of the economy as a whole. CCT programs thus provide a multi-dimensional benefit for participating families. So although this paper focuses narrowly on the health and nutrition impact of the programs, the impact of CCT on poverty, inequality and schooling are also critical to consider when assessing the cost-benefit of the intervention.

CCT programs are spreading rapidly throughout the developing world. Since 1997, seven countries in Latin America and the Caribbean have implemented and evaluated CCT programs with health and nutrition components. These are the subjects of this paper and include Brazil's Bolsa Alimentacao/Bolsa Familia, Colombia's Familias en Acci3n, Ecuador's Bono de Desarrollo Humano, Honduras' PRAF, Jamaica's PATH, Mexico's Progreso/Oportunidades, and Nicaragua's Red de Protecci3n Social. Others are in the process of development in Argentina, Chile, Costa Rica, El Salvador, Panama and Paraguay, and around the world. While evaluation results are encouraging, features of program design are remarkably similar country to country, generating concerns that local realities are not sufficiently taken into account and highlighting the need for better ex ante analysis of the underlying development problem to be solved by the intervention.

This paper will critically analyze the program effect model with respect to health and nutrition outcomes, present the health and nutrition evaluation results to date, draw some conclusions regarding the use of CCT programs to improve health and nutrition status

and promote healthy behaviors, and suggest future directions for the use of CCT programs as an incentive to promote service use and healthy feeding practices. The existing literature well describes and analyzes the design and implementation features of CCT programs (Handa and Davis 2006; Rawlings and Rubio 2003). In recognition of these efforts, this paper will make reference to features of the programs when relevant for the discussion, and annex A includes a brief overview of these features. An annex on the methodological aspects of evaluation is also included as a reference. We also note that this paper focuses narrowly on the health and nutrition impact of the programs.

The remainder of the paper is structured as follows. Section 2 highlights the basic design features of CCT programs, section 3 describes the key demand and supply-side factors that determine use and effectiveness of health care services, and section 4 outlines nine key assumptions that implicitly underlie the program effect model and current evaluation efforts with regards to CCT programs, reviewing and critically examining the evidence available about the plausibility of these assumptions. Section 5 concludes with recommendations for how best to move forward with CCT programs for improving health and nutrition as well as using the process of evaluation to improve design.

II. BASIC DESIGN FEATURES OF CCT PROGRAMS

The basic design features of CCT programs described in this section are: (i) the design and structure of the conditions; (ii) the design and structure of transfers; (iii) eligibility for and targeting of transfers; (iv) program costs; (v) operational arrangements; (vi) supply conditions; and (vii) impact evaluation.

Design and structure of conditions. In order to receive CCT cash transfers month-to-month, beneficiary households must comply with a set of conditions related to the use of preventive health services and attendance to health education sessions. There are no explicit nutrition conditions; rather this outcome was expected to result from the combination of the increased income from the transfers and the nutrition knowledge provided by the education and training components of the programs. Table 1 summarizes the health conditionalities in the 6 programs under study.

Table 1
Health conditionality in CCT programs

Program	Condition
Brazil - Bolsa Familia	(i) Updated immunization cards for children 0-6 years old (ii) Pregnant and breastfeeding women for “regular” visits to health centers (iii) Children 0-15 years old for “regular” visits to health centers
Colombia – Familias en Accion (FA)	(i) Children aged 0-4 years attending growth monitoring visits according to a MOH protocol (6/year for age 0-1, 2/year for ages 1-3, 1/year for age 3-4). (ii) Mothers’ attendance to bimonthly health education workshops
Honduras -- PRAF	(i) Children attending growth monitoring visits according to MOH protocol (ii) Pregnant women for at least four pre-natal care visits
Jamaica -- PATH	(i) Children aged 0-6 attend check ups every two months during ages 0-1 and twice a year thereafter.

Program	Condition
Mexico – Progres/ Oportunidades	(i) Children aged 0-23 months fully immunized and attending growth monitoring visits every two months ¹ (ii) Children aged 24-60 months attending growth monitoring visits every four months (iii) Pregnant women for at least four pre-natal care visits (iv) Breastfeeding women for at least two post-partum care visits (v) Other family members once a year for physical check ups (vi) Adult family members to attend health talks – female heads of household every two months; other adults once a year
Nicaragua -- RPS	(i) Mothers' attendance to bimonthly health education workshops (ii) Up to date vaccination scheme for children aged 0-1 year old (not enforced due to supply failures) (iii) Children under 2 years old attending monthly growth monitoring/well-baby visits (iv) Children between 2 and 5 years old attending bi-monthly medical check-ups

Design and structure of transfers. Transfer sizes related to health and nutrition conditions are generally lump sum. There have been a number of different approaches to the design of the transfer. In Colombia, Jamaica and Mexico, the amount of the health/nutrition transfer represents the difference between the consumption of an average extreme poor household and the food poverty line, with some variations (i.e., the per capita indigence gap for children under 6 only). The goal of this type of transfer is to – on average -- move households living in indigence to a minimum consumption level. In addition to a basic needs and political issue, minimum consumption was seen as a pre-requisite to human capital investment. Another approach, used in Honduras, is to base the transfer amount on the opportunity costs of accessing health care, which has resulted in a lower average monthly transfer. For the purposes of the present paper, it is important to note that the health and nutrition grant's lump sum structure favors smaller families, which may affect program effects, and that it is the combined amount of the transfer, representing both schooling and health/nutrition subsidies, that influence the results achieved in the evaluated programs. On average, payment amounts range from 10 to 25 percent of total consumption among beneficiary households.

Table 2
Payment amounts in current USD²

Program	Monthly Monetary Benefit	Average Monthly Transfer	Average Transfer as a % Poverty Line	% Pre-Transfer Household Consumption
Brazil – Bolsa Familia	\$18/HH; \$5/child up to 3 children	\$24	12%	
Colombia – Familias (FA)	\$20/HH; \$6/child primary age; \$12/child secondary age	\$50	Not reported	30%
Honduras – PRAF	\$4/HH;\$5/child	\$17	8%	10%
Jamaica – PATH	\$9/eligible HH member (child, elderly, disabled)	\$45	16%	20%
Mexico – Progres/ Oportunidades	\$13/HH;\$8-17/child primary;\$25-32/child secondary; one time grant \$12-22/child for school supplies	\$20	23%	25%
Nicaragua – RPS	\$18/HH; additional \$9/HH with a school-aged child; \$20/year/child for supplies	\$25	18%	20%

¹ Visits included measurements, nutrition supplements equivalent to 100% RDA for micronutrients and 20% protein, education for parents on nutrition health and hygiene.

² This table is reprinted with kind permission from Handa and Davis (2006).

Eligibility for and targeting of transfers. CCT programs are directed to poor families; in most cases, families with children. A central feature of CCT programs is the use of explicit targeting strategies to determine eligibility for benefits.

PRAF and RPS apply geographical targeting strategies only. Poor localities are identified using an index of well-being usually constructed on census and survey information, program localities are selected randomly up to a budget constraint and all households within the selected localities are eligible to enroll in the program. Progreso/Oportunidades, FA and Bolsa Familia apply a first round of geographic targeting of localities, followed by the application of proxy means testing or direct income testing to identify individual households eligible to participate in the programs. Other conditions for the participation of localities are sometimes established; for example, the FA program requires that participating municipalities have a bank available within a given geographic reference area as well as sufficient supply of health and education services to be available to meet expected increases in demand.

With the exception of Progreso/Oportunidades, CCT programs restrict entry to poor households with young children, school-aged children and pregnant women; all other households are excluded. Categories of eligible children vary depending on (i) the country's nutrition strategy that sets certain age groups as targets (priority has been shifting from 0-5 to 0-3 years old, six year olds are rarely included); (ii) the official starting age for school; and (iii) if the objective of the program is primarily human capital creation or consumption smoothing (in the latter case the program should not exclude 3-6 year olds). In PRAF, for example, children under age three and children aged 6-12 who have not yet completed fourth grade of primary school are eligible; households with children between ages three and six and no children in other age groups are excluded. In RPS, young children up to age five are eligible and school aged children ages 7-13 are eligible but households with children age six and no children in other age groups are excluded.

Coady, Grosh and Hoddinott (2004) have reviewed the efficiency of different targeting strategies, finding that although (proxy) means tests' performance varies significantly according to implementation capacity, these instruments display the best results in terms of errors of inclusion and exclusion and can be cost-effective in specific settings. While the Mexican targeting strategy reportedly generated social conflicts related to targeting individual households in majority poor communities (Adato et al 2000; Coady 2001) and there is some evidence of beneficiary gaming of the targeting system, the efficiency gains from household targeting appear to be substantial, particularly as programs expand into more economically heterogeneous settings (i.e., urban areas). For example, an IFPRI study for Nicaragua's RPS (2002) shows that the addition of household targeting to geographic targeting could have increased the share of cash benefits reaching the poor from around 88 percent to 95 percent while leakage rates would decline from fifteen to five percent. Handa and Davis (2006) provide an excellent overview of the targeting practices of different programs and their implications for selection.

As with all targeting instruments, the use of individual household targeting also has limitations, which may be important if the program is to serve as a safety net. Selecting

households based on determinants of poverty emerging from a principal components analysis has tended to prioritize families with small children while excluding poor households without small children or with elderly heads (the Colombia program excludes households without small children from eligibility for the health and nutrition grant in any case). Problems have also been detected around the use of point-specific eligibility cut offs, which somewhat arbitrarily excludes households that may differ from eligible households in insignificant ways, leading to the adverse community effects observed in the Progresa evaluation. Finally, while individual household targeting can improve program efficiency in more heterogeneous settings, in small, rural and highly marginalized communities, the Mexican experience suggests that geographic targeting alone is sufficient and that the more precise targeting strategy employed by the program is not cost-effective.³

The process of beneficiary re-qualification for the programs varies from country to country. Mexico applies the proxy means test questionnaire to existing and potential beneficiaries on a periodic basis. New families may be incorporated and existing beneficiaries who are no longer poor – as defined by the means test – are determined ineligible. Colombia and other countries initially set limited terms of eligibility, generally associated with the expected duration of available financing, but have since modified the strategy to one more similar to Mexico. In the case of countries that provide benefits to individuals within poor households rather than the household as a whole, the socio-demographic characteristics of the beneficiary household (i.e., presence of children in eligible ages) determine continued eligibility.

Most CCT programs have been scaled up gradually and, with the exception of Oportunidades as of 2005, are programs that do not reach the universe of its intended beneficiaries. This has been due to a desire to test operational procedures and measure the results of the program using impact evaluation methods, but is also attributable to budget limitations.

Table 3
CCT beneficiary characteristics

Program	Number of Beneficiary HH	% Poor	Year of Start
Brazil - Bolsa Familia	~2 – 5 million (2004) ⁴	Not reported	2003
Colombia -- FA	362,000 (2005)	87% poor; 44% indigent	2001
Honduras -- PRAF	30,000 (2006)	5% extreme poor	2000
Jamaica – PATH	180,000 (2005)	Not available	2001
Mexico -- Progresa/Oportunidades	5,000,000 (2006)	100% extreme poor	1997
Nicaragua – RPS	30,000 (2004) – now closed	Not reported	2000

Program costs. Little has been done on the comparative costs of programs; the exception is Caldés et al 2004 which reviews Progresa, PRAF and RPS. Handa and Davis 2006

³ It should be noted, however, that there are other motivations for the use of formula-based individual (proxy) means testing such as the reputation of the program for impartiality and transparency. The Mexico targeting strategy was based on such motivations, although the perception of beneficiaries differed substantially (reword: technical people believed in its transparency and impartiality even while beneficiaries complained that they could not understand how eligibility was determined; both are probably important).

⁴ Bolsa Familia’s administrative records indicate 5.04 million beneficiaries while household survey data shows approximately 2.09 million (Soares et al 2006). The discrepancy is mostly unexplained though attributed to a period of transition from Bolsa Escola/Alimentacao to Bolsa Familia.

provide the most recent estimates of program costs, noting that these amount to less than 1% of GDP in most cases.

Operational arrangements. CCT programs have tended to be implemented by specially created entities linked directly to Presidential offices or other semi-autonomous project executing units. As such, they are institutionally separate from local governments and “line ministries” such as the Ministry of Health. This seems to have contributed to the rapid pace of execution observed in most programs but has also generated institutional and bureaucratic friction among the implementing partners.

In general, the project cycle of a CCT program involves a rough sequence of activities including: (i) selection of program areas (geographic targeting) and coordination with health and education sectors; (ii) identification of beneficiary households (household targeting); (iii) enrollment of beneficiaries, generally involving beneficiary meetings in each community to inform participants of their rights and responsibilities under the program; (iv) organization of supply responses in advance of delivery of demand transfers and delivery of supply transfers where relevant (Honduras, Nicaragua); (v) verification of conditionality involving the distribution, collection and processing of clinic and school attendance records; (vi) delivery of demand transfers, calculating transfers based on compliance levels, informing beneficiaries about scheduled transfers and ensuring that the disbursement and payment of transfers through banks or post offices is conducted in a timely and orderly manner; and (vii) internal monitoring and evaluation, including supervision, spot checks, audits, etc.

Supply conditions. In some countries, minimum supply conditions had to be met prior to implementation of demand-side piece. In Colombia, this took the form of a minimum provider/infrastructure to beneficiary ratio and the availability of “space” to produce more visits with respect to a standard ratio. In Mexico, minimum distances to facilities were established. In Honduras and Nicaragua, however, CCT programs built in supply-side strengthening intended to respond to the pressures associated with increased demand for services and the possibility that quality might decline as a result of higher productivity. This strengthening took the form of contracting private providers for the provision of a package of services and health education talks in Nicaragua (\$160/beneficiary/year) and support to a quality improvement and basic package through public health centers in Honduras (\$6,000/health center/year). In both cases, there were substantial delays associated with the implementation of these components, and these delays were greater than those associated with designing and launching an entirely new CCT program (a commentary on the relative efficiency of MOH). The supply-side contracts in the Nicaragua program included pay for performance (P4P) incentives.

Impact Evaluations. Conditional cash transfer programs are unique in their use of multi-sectoral interventions and in the decentralized nature of implementation. Initially, little was known about how well local governments and institutions would be able to implement such programs and how well the programs would perform in improving child health and education outcomes. Thus, each of the six programs discussed in this paper prioritized conducting a rigorous impact evaluation to measure the impacts on key outcomes. These impact evaluations, while strongly promoted by researchers and experts involved in designing the programs, were also accepted by governments as a way to

provide evidence of success in improving child outcomes and to gather public support. Annex A discusses the impact evaluations conducted for each of the programs.

The evaluation designs have varied in their ability to produce estimates of impacts that can be generalized to the full beneficiary populations or to similar programs. Generalizations of the impact results are limited due to issues of sample selection and the fact that instead of evaluating each component separately, the evaluations focused on the impacts of the package of multiple interventions. Granted, separating the effects of the different components poses significant challenges, but attempts in this regard were limited at the outset by the choices and/or design of instruments and the general approach to the evaluation.

The evaluations typically seek to confirm the existence and measure the size of expected impacts, identify unanticipated effects, understand beneficiary and stakeholder perceptions of the program and verify that program benefits are delivered in a cost-effective manner. To achieve these objectives, the CCT evaluations have experimental (Honduras, Mexico, Nicaragua) or quasi-experimental (Colombia, Ecuador, Jamaica) designs, with repeated observations from large samples of households in treatment and control groups via specially designed surveys conducted before and after program implementation. School and health center surveys or administrative data on supply for program intervention areas are available for Colombia, Honduras and Mexico, and four countries included a qualitative evaluation on some aspects of the program's operation and effects.

In general, the experimentally designed evaluations represent the best attempt to measure the impact of CCT programs as sample selection bias is limited through random assignment. Quasi-experimental designs must overcome the selection issue but when done carefully can limit the bias that results and there are many econometric strategies available to overcome the design limitations. Less often discussed is the portability of results both within the country and across countries. It is rare that an evaluation is designed such that impacts can be estimated for the full population of intended beneficiaries. The evaluations of Oportunidades, PRAF, RPS and PATH all suffer in this way as the evaluation samples (or pilot phases on which evaluations were conducted) are not representative of the full or potential beneficiary population. For example, the evaluations of PRAF and RPS were conducted on pilot phases, which were implemented on a select portion of the country. In the case of PRAF, the pilot beneficiary group represented the most impoverished communities in the country. In the case of RPS, the pilot phase was not offered to the poorest localities, but was still not representative of all potential beneficiaries. Table 4 summarizes the characteristics of the published evaluation data by country; results are only reported for a sub-set of programs that have completed and reported on at least one post-program round of data collection: Colombia, Honduras, Mexico and Nicaragua.

Table 4
Summary of Evaluation Data by Country

	Country			
	Mexico	Honduras	Nicaragua	Colombia
Evaluation Period	Rural: 1997 –1999 Urban: 2002 - 2003	2000 – 2002	2000 – 2002	2002 - 2003
Targeting level (method)	locality and household (marginality index)	Locality (priority index)	municipality (70 with lowest haz of first graders)	locality (meet 4 criteria) and household (SISBEN 1)
Evaluation Design	Rural: Experimental Urban: Quasi-experimental	Experimental	Experimental	Quasi-experimental
Main Evaluation Methodology	Difference-in-differences; Propensity Score Matching (urban)	Difference-in-differences	Difference-in-differences	Difference-in-differences; Propensity Score Matching
Sample size (panel) Households (individuals)	Rural: ~22,000 (~110,000) Urban: 17,201 (76,002)	5,096 (25,777)	1,396 (~8,500)	10,742 (64,500)

In addition to impact evaluations, some programs (OPORTUNIDADES, RPS, FA, PATH) included evaluations of the targeting method and program process and implementation in their initial program design. In Latin America, the design of conditional cash transfer programs came at a time when it was widely recognized that public transfers were often poorly targeted to poor households and individuals and frequently captured by small subgroups with strong political ties. Therefore, evaluations of the targeting methods, when conducted, were determined to be critical for ensuring support from both the public and international donor organizations. This was most notably the case in Mexico, where OPORTUNIDADES marked the beginning of sweeping reforms in the design and delivery of public transfers (Levy and Rodriguez 2004). Evaluations of the programs' procedures and implementation provide country governments with valuable information about how to improve program operations and in many cases could help explain the outcomes observed through the impact evaluations. In some cases, the lessons learned could be generalized to other countries and programs and added to the world's knowledge about how to effectively design and implement conditional cash transfer programs.

III. UNDERSTANDING THE DEMAND AND SUPPLY-SIDE FACTORS THAT DETERMINE UTILIZATION AND EFFECTIVENESS OF HEALTH CARE SERVICES.

Understanding the causal pathways that underpin an intervention is critical to evaluating how it works and what should be modified to improve effectiveness. This section will review the factors that influence the demand for and supply of health services.

Demand-side factors. The level of use of health services is determined by a combination of demand- and supply-side factors. Illness reduces market and non-market productivity, the total amount of time available for production, as well as individual well being. To avoid or minimize these unwelcome experiences, individuals will tend to invest in their health in order to produce health, or at least restore part of it after an illness. These ideas were presented in an article by Selma Mushkin (1962), and formalized by Michael Grossman (1972). The extent to which the desire to invest in ones health is reflected in the demand for health care depends on whether an individual identifies illness and is willing and able to seek appropriate health care (Ensor and Cooper, 2004). The identification of illness may depend both on the type of illness, as well as on the level of knowledge of the individual. Willingness to seek care is affected both by knowledge and perceptions (social norms), as well as by direct and opportunity costs of seeking treatment, household income, along with the costs, quality and availability of substitute products and services and intra-household decision-making. The latter may possibly include differences in preferences for males over females or for working adults over children and the elderly.

Supply-side factors. As noted in Eichler (2006), the supply of health care services is determined by a combination of structural inputs (staff time, infrastructure, drugs and supplies, land, etc.) and the processes that transform these inputs into outputs (such as the available technology and the management capability of the provider) that ultimately lead to improved health outcomes. Central to the transformation of inputs into outcomes is the behavior of the health care provider, both at the individual and institutional level. The objective functions of individual providers could include, in addition to the desire to cure patients, the desire to make money and the desire for leisure time. Thus, deficiencies in the quality of care, which tend to impede the ability of health systems to improve health outcomes for the poor are associated first and foremost, with inappropriate incentives for providers along with insufficient resources, organizational rigidities and lack of knowledge.

The next section will critically examine the CCT program effect model in health by discussing the assumptions behind the CCT programs. Selected impact evaluation results will be presented as they relate to each assumption, and will feed into the discussion on whether the implicit assumptions made when designing the CCT programs tend to hold. For a more comprehensive discussion of the evaluation results with respect to utilization of health services, health knowledge, attitudes and practice, consumption, supply of health services, quality of health services, vaccination rates, nutritional status, morbidity, mortality, and fertility, please refer to annex B.

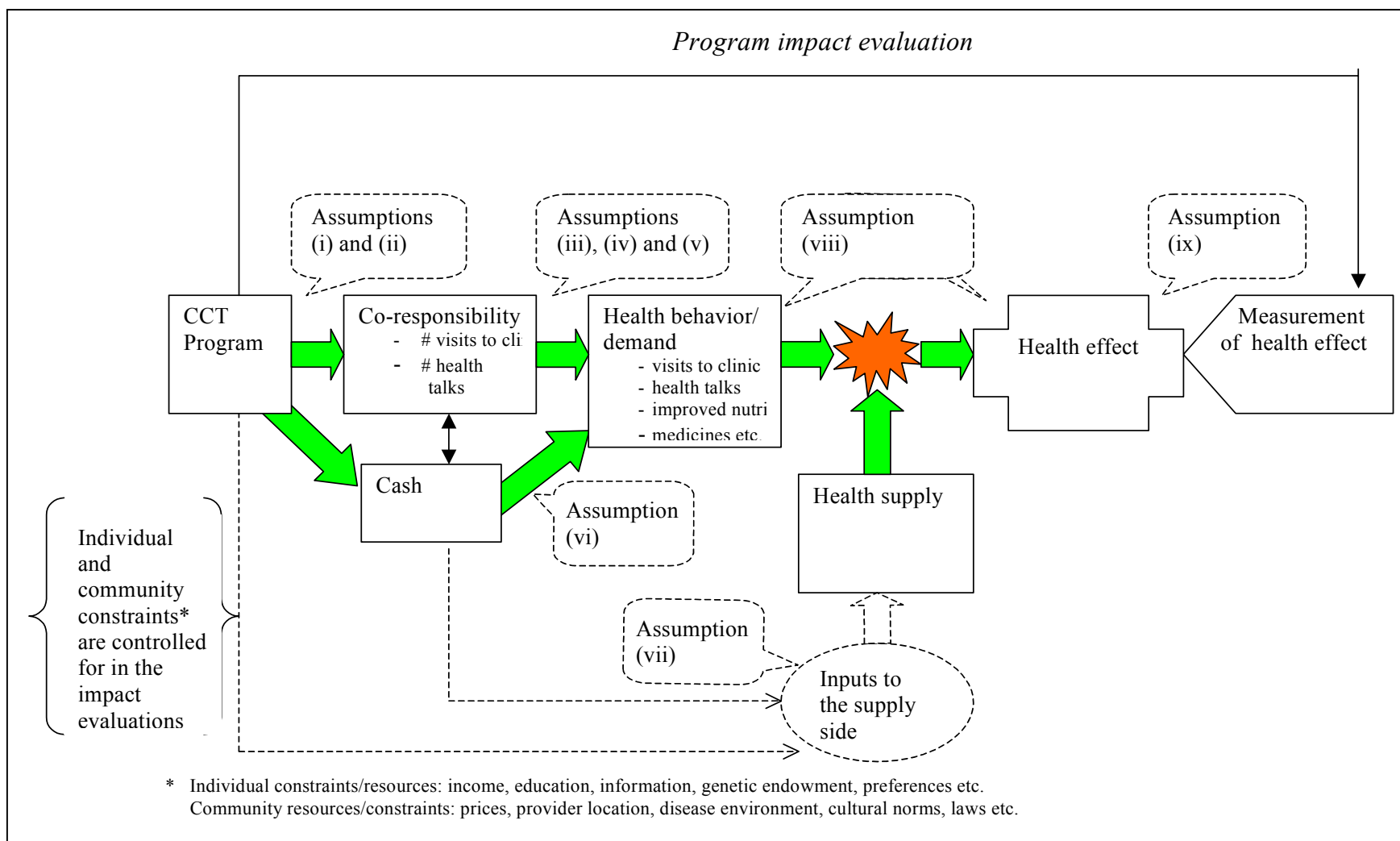
IV. ASSUMPTIONS UNDERLYING THE CCT PROGRAM EFFECT MODEL AND IMPACT EVALUATION METHODOLOGY.

Conditional cash transfer programs, and the documents that describe them, do not tend to make explicit the assumptions underlying their design, particularly as regards the health conditionality. However, from the description of the situation prior to the programs, as well as the design of the interventions, the program effect model can be made explicit.

The typical CCT program in Latin America has tended to include little in terms of direct interventions on the supply-side (Nicaragua is the exception), but does involve inter-sectoral agreements with ministries of health to provide services to program beneficiaries (in some countries this commitment is made more explicit by actually tagging the budget-line going to the health sector, so that it is earmarked for CCT program beneficiaries). The CCT program approach is to stimulate demand for health services and educational health talks by transferring cash to the mothers conditional on attendance to clinics for consultations and health education talks.

Based on this design, a number of implicit assumptions are made. First, that utilization of health services is sub-optimal among the poor. Second, that an increase in utilization of (public) health care services will improve health status. Third, that poor women lack sufficient health knowledge and fourth, that a transfer of information to them will induce behavior changes. The fifth and sixth assumptions are also related: that imposing conditions and that monitoring compliance are necessary to increase utilization of services. The specification of the woman as the beneficiary assumes that the intra-household allocation of the additional resources depends on who is named as the recipient. The eighth assumption identified is that the existing supply of services is sufficient or will increase following increases in demand. Finally, the evaluation methods chosen implicitly assume that the areas impacted by the program are those that are measured and that data is collected appropriately to be able to measure the impact. Each of these assumptions is discussed in turn below, and the program effect model is represented schematically in Figure 1.

Figure 1
Program effect conceptual framework



4.1 Sub-utilization of existing health services by the poor.

The idea to condition payments on regular check-ups at a health clinic is based on the assumption that the poor make insufficient use of preventive health services. However, insufficient use is not well defined and therefore the optimum/goal is also not clear. In economic terms, a private optimum amount of use might be determined as the goal, defined as the amount of use that would be observed if poor families had full information about the costs and benefits of preventive health care. A social optimum amount of use might also be used as a goal, understood as the amount of use necessary to achieve a given level of positive externalities that one person's preventive care has on the rest of society.

This assumption was generally explored during the design of CCT programs. Baseline health and nutrition documentation in CCT countries indicated significant inequalities in the use and fiscal impact of health care by socioeconomic strata. Poor and rural households were much less likely to identify illness and seek care when sick than their better off counterparts; their out of pocket spending on health was more significant as a proportion of their total expenditures compared to that of the wealthy and direct and indirect costs associated with care-seeking were frequently cited as a reason for non-use of services. Even in systems with strongly progressive public spending on health, the poor displayed lower use rates. In Jamaica, for example, only seventy five percent of 0-6 year olds had visited a health practitioner in the 12 months prior to the program start date. The proportion was less than sixty percent with respect to the use of preventive care (Mathematica 2004).

Impact evaluations do indicate that the programs can increase utilization. While effects on utilization varied (see Table 1) there is also some evidence that the program caused a substitution of providers. Gutiérrez et al (2005) report that the use of public health services in rural Mexico increased in the same proportion that private ones decreased, suggesting that beneficiary families in Oportunidades may have changed their health care seeking behavior by substituting public for private services.

However, there was little ex ante analysis conducted to directly test the proposition that the inequities in health and nutrition outcomes and outputs were primarily due to demand-side factors relative to supply-side factors. As Handa and Davis (2006) note, "the common observation that poor children attend ... health check-ups less frequently is not sufficient evidence to conclude that a demand-side intervention will solve the problem since income is highly correlated with access to and quality of ... health care making it just as plausible that differences in outcomes are driven by the unequal distribution of quality services." And even if both supply and demand side factors are shown to be important, the question remains as to which is the more cost-effective option for the government to pursue.

Two studies have attempted to estimate the relative cost-effectiveness of supply versus demand-side interventions in improving schooling enrollment in developing countries (Coady and Parker 2002; Handa 2002). However, the data used in these studies, which did not include any important changes in the supply side during the periods considered, did not allow for a rigorous investigation into this policy question. While no similar

health or nutrition study has been undertaken to date, Handa and Davis (2006) notes that health care, particularly preventive health care, differs from education in that information asymmetries are more acute and that poor households may be less likely to seek care for this reason, which would justify government intervention to correct this market failure.⁵

Table 5
Program Effects on Utilization of Health Services⁶

Indicator/Country	Mexico (1998-2000) ^{7 8,9}	Honduras (2000-2002) ¹⁰	Nicaragua (2000-02) ¹¹	Colombia (2001-2003) ¹²
Public clinic visits for children (reference period)	<i>Rural</i> 0-2 years old 1.5% decrease 3-5 yrs old No impact (last 6 mos.)	0-3 yrs old 20.2% increase (last 30 days)	0-3 yrs old 11% increase (last 6 mos.)	0-2 yrs old 30% increase 2-4 yrs old 50% increase (completed age appropriate visits)
Pre-natal care visits (number of visits; details)	<i>Rural</i> No impact <i>Urban</i> ¹³ 6.12% increase (4 or more; Kessner index ¹⁴)	18.7% increase (5 or more; last pregnancy)	Not reported	Not reported

4.2 Utilization of (public) health care services will improve health status.

By conditioning on the utilization of preventive health controls, primarily in public sector clinics, the CCT programs are clearly assuming that these actions will improve health status. This is a multi-layered assumption that relates both to the quality of the services given at the public clinics, as well as to the quality and effectiveness of substitute products and services. If utilization is low due to poor quality, increasing utilization

⁵ Handa and Davis (2006): “The demand for quality health care is difficult to model because it is hard to measure (and control for) the exogenous price of different alternatives, but there is evidence that both quality and access are also important determinants of utilization (Akin et al 1995; Jensen & Stewart 2000).”

⁶ As is noted at the beginning of this section, each evaluation’s sample design determined the type of analysis that could be conducted on the data. In general, difference-in-differences estimates are used to report the impacts.

⁷ In Mexico, utilization was not included in the baseline survey, requiring first difference estimates of impact. Estimates of impact on utilization in Mexico were also conducted using difference-in-difference with clinic-level administrative data, available before and after the program. The administrative data confirmed an increase in utilization in program communities, but these utilization statistics were not disaggregated by age group.

⁸ Public clinic utilization for children, Gertler and Boyce (2001).

⁹ Prenatal care visits, Prado et al (2004).

¹⁰ Impact refers to the household package only (i.e., demand-side subsidies only) as reported in Morris et al (2004b).

¹¹ Maluccio and Flores (2004).

¹² Unión Temporal IFS-Econometria S.A.-SEI (2004).

¹³ Reference period is 2002 – 2003 for urban areas.

¹⁴ Index of care received determined by the number of visits and the timing of each visit; does not include any measure of quality of care received at each visit (Prado et al 2004).

without improving quality could have the opposite effect of that intended. We review here the impacts found on nutritional status, morbidity and mortality. While improvements in health status are observed in some programs, none of these impacts can be tied directly to the use of health care services.

Unlike the unambiguously positive results with respect to food consumption, nutritional status outcome measures show a mixed result. Leaving aside the Honduras program, all of the programs show a significant reduction in stunting. Results for the proportion underweight are less consistent; Nicaragua RPS shows a large and significant decline, while the Colombia program shows impact only in rural areas for 3–7 year-olds. No effects on anemia are observed. Table 2 summarizes program effects on nutritional status.

Table 6
Program effects on child nutritional status

Program	Mexico (1997 - 1999) ¹⁵ (1997 - 2003) ¹⁶	Honduras (2000-2002)	Nicaragua ¹⁷ (2000-2002)	Colombia ¹⁸ (2001-2003)
Proportion stunted; height-for-age (haz) ≤ -2.0	<u>1997 - 1999</u> 12 - 36 months decrease (coefficient for logit estimate reported) <u>1997 - 2003</u> 24 - 72 months 29% decrease (girls) 11% decrease (boys)	0 - 4 yrs old no impact	0 - 4 yrs old 5.5% point decrease	0 - 2 yrs old 6.9% points decrease 2 - 7 yrs old no impact
Proportion underweight; weight-for-age (waz) ≤ -2.0	no impact	0 - 4 yrs old no impact	0 - 4 yrs old 6% point decrease	Rural 0-3 yrs old no impact 3 - 7 yrs old 3.4 % points decrease Urban no impact

Morbidity may increase or decrease as a result of the CCT intervention. On the one hand, greater use of preventive care and higher levels of health knowledge may lead to fewer episodes of illness. On the other hand, greater levels of health knowledge and more frequent visits to health centers may increase the probability that mothers more frequently and/or accurately diagnose illnesses and seek care when it is required. Moreover, improved health may lead to increased symptoms of morbidity (such as respiratory infections) which cause an increase in the demand for curative care (Oppenheimer 2001). These hypotheses have not been much explored by the evaluations, but where it has been measured (mainly in Mexico), CCT appear to decrease the incidence and prevalence of morbidity.

¹⁵ Behrman and Hoddinott (2000), data collected by Instituto Nacional de Salud Publica (INSP) 1998 and 1999 that was then matched to the ENCEL evaluation data.

¹⁶ Gertler and Fernald (2004).

¹⁷ Maluccio and Flores (2004).

¹⁸ Unión Temporal IFS-Econometria S.A.-SEI (2004).

The Mexico evaluation found that the impact of the program on the probability of child illness is negative and statistically significant for all age groups, but not until a child had been receiving benefits for at least a 12-month period (Gertler and Boyce 2001). In rural areas, a small decrease in sick days for only the productive age population is observed, while a larger effect and wider age range is seen in urban areas. Overall, the number of days lost to illness decreased by 20% among beneficiary families (Gutierrez et al. 2005).

An Oportunidades study of the effect of the program on indigenous populations found that the program decreased illness rates from 0.2% to 3.5%, with the greatest effects observed among children under three years old (Quiñones 2006). The average program effect estimations differ substantially from Gertler's findings, most likely due to differences in sampling size and approach used. Positive program effects observed for indigenous beneficiaries were similar to those of non-indigenous beneficiaries. This finding could be considered a weakness of the program, as is argued by Quiñones, given the implicit assumption that the average effect of the program should be larger on poorer and more marginalized groups. On the other hand, it is remarkable that the benefits experienced are equivalent for both groups when one takes into account evidence that indigenous groups have a more difficult time complying with CCT conditions due to language and cultural barriers.

With respect to chronic disease-related morbidity, the required regular check-ups and participation in health talks may have a positive effect on household and social norms related food intake and activity. However, the income transfer to the household could make some behaviors that increase the risk of chronic diseases more affordable (e.g. junk foods, soft drinks, alcohol). The single study on this issue from Mexico finds encouraging results. Fernald et al. (2004) find high baseline prevalence of obesity (twenty percent), hypertension (thirty nine percent), and diabetes (nineteen percent) among the rural poor in Mexico and that Oportunidades participation significantly reduces the prevalence of all except diabetes. Symptoms of hypertension and diabetes are also significantly reduced via program participation.

Mortality was not measured directly by the evaluations; only the Mexican program uses administrative data to analyze program effects. Hernández et al. (2004) examine the impact of Oportunidades on maternal and infant mortality using data for the period 1995-2002 from the Ministry of Health (SSA) and the National Institute of Statistics, Geography and IT (INEGI). For the entire period, maternal mortality was eleven percent lower in the municipalities with at least one locality incorporated in the Oportunidades program compared to those without any localities incorporated. For the entire period (1997–2003), infant mortality was two percent lower in the municipalities incorporated in Oportunidades compared to the non-incorporated ones. Estimates of absolute numbers indicate that in this period an average of 340 infant deaths per year have been avoided thanks to the program. The impact of Oportunidades on infant mortality at the municipal level increases relative to the proportion of the population incorporated in the program.

4.3 **Insufficient health education and knowledge among poor women.**

The inclusion of educational health talks, mainly targeted towards the women of beneficiary households, as part of conditionality implies an assumption that it is not only insufficient income that is the reason for inadequate health investments in the household, but also a lack of information and education. However, direct measurement of health knowledge and attitudes has generally not been a component of the evaluations (or if included in questionnaires, has not been reported in evaluation reports or papers as in the case of Colombia), in spite of the inclusion of health education components in all programs.¹⁹ As knowledge, attitudes and practices are posited to be critical elements in health outcome changes, understanding the evolution of these factors is critical to explaining the health results observed.

4.4 **Information induces behavior change.**

Perhaps not by design, but by default, the educational health talks have tended to expose beneficiaries to health information passively. Expecting that this type of intervention will have an effect on the health seeking and health inducing behavior of households implies an assumption that information in and of itself will induce behavior change. There are a few exceptions to the lack of study of the impacts of providing health and nutrition information to beneficiaries on health knowledge and behaviors.

In addition to the increase in food consumption resulting from the cash transfers (see Table 7), Progresa evaluators found an increase in dietary quality and calorie consumption. After controlling for the income effect associated with increased calorie consumption, the increase in consumption of more diverse, high nutritional quality foods, such as fruits, vegetables and animal products, indicates that there may be an effect of the nutritional education provided through health education talks known as “pláticas” (Hoddinott et al 2000). This effect did not vary systematically by socio-economic status of the beneficiary families. Duarte et al (2004) attempts to isolate the effects of the Oportunidades health talks on smoking and alcohol consumption in rural areas, an analysis that is also interesting given that a potential income effect of the transfer is to increase consumption of these goods, finding that youth consumed less alcohol and more cigarettes than control groups. Moreover, Prado et al (2004) report an increased knowledge of family planning methods in both urban and rural areas. In addition, in rural areas use of modern family planning methods is higher in the intervention than in the control group. Breastfeeding initiation and duration was included in many questionnaires, but few reported on results. Surprisingly then, the lecture-style health talks accompanied by peer to peer learning through the “community mothers” seem to have had positive effects in Mexico in some areas. Moreover, Davis et al (2002) find that the impact of Procampo (a transfer program designed to compensate farmers for lost crop income after NAFTA) on dietary diversity was greater when the family was also a Progresa beneficiary, indicating that the information transfer positively influenced health-related behaviors.

¹⁹ The Colombian program evaluation survey included questions pertaining to the knowledge of appropriate care for diarrhea. However, this information was not studied for the evaluation.

Table 7
Program Effects on Household Consumption²⁰

Indicator	Mexico²¹ (1997 – 1999)	Honduras²² (2000 – 2002)	Nicaragua²³ (2000 – 2002)	Colombia²⁴ (2001-2003)
Mean per capita Food Consumption/Expenditures	10.6% increase	no impact	21% increase	20.4% increase urban areas 22.5% increase rural areas (household total)
Mean per capita Total Consumption/Expenditures	not reported	no impact	13% increase	13.9% increase urban areas 16.9% increase rural areas (household total)

Evaluators found that the Colombia FA program increased the time that children are breastfed by 1.44 months in urban areas and 0.84 months in rural areas. In addition to increasing per capital consumption (Table A3), the program also improved the quality of food consumed by children, increasing the average number of days per week in which various proteins (between 0.25 days and 1.09 days), grains (about 0.45 days) and fruits and vegetables (between 0.45 days and 1.23 days) are consumed. In addition, the program increased overall household consumption of high quality foods - proteins, tubers, cereals, fruits and vegetables and fats and oils in urban areas and proteins, cereals and fats and oils in rural areas. More evidence of CCT program impact on improving healthy practices comes from Maluccio and Flores (2004), who find that RPS not only increased overall food and total consumption, it also had a beneficial effect on dietary diversity in terms of both the number and quality of foods purchased (Table A3).

4.5 Conditioning necessary to induce desired levels of utilization.

In order to receive CCT cash transfers month-to-month, beneficiary households must comply with a set of conditions related to the use of preventive health services and attendance to health education sessions. By conditioning the transfer on certain types of desired behavior, CCT programs are implicitly assuming that the relaxation of the budget constraint brought about by the monetary transfer will not be sufficient to induce major changes in human capital investment. Instead, a condition that transforms the income effect into a price effect, for example, by requiring regular health check-ups, needs to be in place. However, this assumption may not hold. There may be a level of transfer which

²⁰ Difference in difference estimates are reported for Nicaragua and Colombia and Honduras for food expenditures. Cross-sectional estimates (first difference) are reported for Mexico and total expenditures in Honduras. In the case of the former, no consumption data was collected at baseline; in the case of the latter, there was a difference in the seasonality of the baseline survey across the control and treatment groups that affected baseline non-food expenditures.

²¹ Hoddinott et al 2000.

²² Impact refers to the household package only (i.e., demand-side subsidies only) as reported in Morris et al 2004.

²³ Maluccio and Flores 2004.

²⁴ Consorcio IFS-Econometria-SEI 2004, estimates reported for difference-in-differences between treatment group without payment (TSP) and control group; converted to percentages from the log-point estimates.

would induce the desired behavior without conditioning. In that case, the relative cost-effectiveness of a conditioned and non-conditioned transfer scheme should be calculated. A major potential source of inefficiency in the CCT programs is paying people for what they were already going to do.²⁵ In order to reduce this efficiency leakage, program designers should ideally model the probability that a given beneficiary will attend the conditioned health services under different transfer scenarios.

Conditionality seems to work, though there are few well documented studies on this subject. A comparative ex post impact evaluation by Davis et al (2002) of the Mexico Progresa program (pre-cursor to Oportunidades) and an unconditioned transfer to poor farmers known as Procampo found that while overall increases in food consumption were comparable between the two programs, the effects on health and schooling were significantly greater under Progresa, suggesting that conditionality did make a difference for total program's effects. However, the transfer recipient is most usually a male in Procampo, while in the CCT program mothers or primary females are the main recipient, which could be the source of differences in outcomes, rather than the conditions.

Although not a comparative study, Agüero et al (2006) finds that an unconditional cash transfer in South Africa, where only the title of the grant (Child Support Grant) implies the expected use of the money, increases nutritional status as measured by height-for-age. Significant impacts are found only when households benefit before the child's second year of life, consistent with nutrition literature that indicates that the first two years of life are most critical for a child's growth (Allen 1993; Martorell 1999) and with impacts found in CCT programs (see Table A4). Thus, the findings indicate that conditionality is not required for a cash transfer program to have an impact. However, there is currently no measure of whether conditions would have improved nutritional status further.

4.6 Only if the conditionality is monitored and compliance enforced will it have an effect.

Program designers have feared that without monitored conditionality and disincentives for non-compliance (docking transfers when conditions are not met), CCT beneficiaries will not comply with program conditions. Two important aspects may counter this effect. The first is that the mere signaling by the authorities (or program officials) that compliance will be monitored passes a message to the beneficiaries of the importance of the activity. Second, the presence of conditions implies that there is a risk of losing the transfers, which may linger for a long time.

While conditions are always stated, compliance is not necessarily monitored. Of the cases where there was effective monitoring, as in Colombia, Mexico, Nicaragua and Jamaica, compliance is generally extremely high, ranging from a high of ninety nine percent of Oportunidades households in Mexico (Behrman and Todd 1999a) to a low of ninety four percent in Jamaica (Mathematica 2004). In cases without monitoring, evidence is

²⁵ See Sadoulet et al (2002) on education.

mixed.²⁶ Morris et al (2004b) find null effects in the Honduran program where conditionality was not enforced, though payments were also irregular and may be responsible for the absence of observed effects. Schady and Araujo (2006), however, find that just the announcement of conditionality, even if not enforced, was sufficient to induce a large and significant change in the behavior of poor households in Ecuador with respect to school attendance.²⁷ However, similar work has not yet been done for health. In addition, there is limited knowledge about how long the mere *threat* of monitoring compliance can substitute for actual compliance.

4.7 Household allocation of the additional (CCT) resources depends on who is the recipient.

CCT programs designate the mother or primary female as the recipient of the transfer under the explicit assumption that they are more likely to invest in the welfare of the children in the household. Note that this implicitly assumes that the transfer recipient also decides on its use. While Davis et al (2002) find that a CCT program (Progresa) in which a female is the recipient leads to larger improvements in health and education than an unconditional transfer for farmers in which the male was most often the recipient (Procampo, intended to compensate for losses due to NAFTA) suggests that the identity of the beneficiary matters. However, this evidence is also confounded by the fact that Progresa imposed conditions while Procampo did not.

4.8 Supply-side of services is in place or will follow demand.

With the exception of CCT programs that include supply side strengthening, such as Nicaragua, most programs assume that existing supply side capacity is sufficient to meet CCT beneficiary demand. This may be based on the common observation that provider productivity in the public sector is low and could be increased if incentives were aligned even if other inputs remain unchanged. In some countries minimum supply conditions had to be met prior to implementation of demand-side piece. In Colombia, this took the form of a minimum ratio of providers and infrastructure to beneficiaries as well as the availability of “space” to produce more visits with respect to a standard ratio. In Mexico, minimum distances to facilities were established.

In Honduras and Nicaragua, however, CCT programs built in supply-side strengthening intended to respond to the pressures associated with increased demand for services and the possibility that quality might decline as a result of higher productivity. This strengthening took the form of contracting private providers for the provision of a package of services and health education talks in Nicaragua (\$160/beneficiary/year) and

²⁶ In Honduras, for example, women were required to deposit a bar-coded, certified attendance slip in an urn for every required visit to a health center, however, no beneficiary was ever suspended for non-compliance (Morris et al 2004a).

²⁷ Schady and Araujo (2006) report effects on enrollment that are two and a half times as large as those observed in Progresa. The author attributes this difference to the much lower baseline level of school attendance and enrollment in Ecuador as opposed to Mexico. A similar phenomenon is observed across other impact evaluations; those countries that start with lower baselines see larger effects, which may indicate the appropriateness of CCT in poorer countries and areas without the need for intensive monitoring of conditionality.

support to a quality improvement and basic package through public health centers in Honduras (\$6,000/health center/year). In both cases, there were substantial delays associated with the implementation of these components, and these delays were greater than those associated with designing and launching an entirely new CCT program (a commentary on the relative efficiency of MOH). The supply-side contracts in the Nicaragua program included pay for performance (P4P) incentives.

It is often assumed that if the problem is on the supply side, then the transfer needs to be made to the supply-side, ideally in the form of an incentive to perform, and if the problem is one of lack of utilization due to a lack of knowledge of the benefits of preventive care or due to deficient information regarding the existence of the services etc. then the transfer should be made to the demand-side, i.e. the beneficiaries. Nevertheless, under the assumption that the beneficiaries are well-informed of the potential benefit from preventive health care, if the problem is one of quality and existence of health care provision, then the beneficiaries can use their additional cash from the monetary transfer to incentivize the supply-side.

In countries that did not purposely boost supply in tandem with the cash transfers, the hope was that governments and providers would increase supply-side inputs if beneficiaries began to demand services and provider accountability. In other words, as a response to conditions, beneficiaries may protest in front of the health center until the public health authorities respond (as was recently the case in Panama as a reaction to the CCT pilot) or use part of their additional cash to encourage providers to work longer hours or obtain drugs and supplies. Whether or not these changes in supply resulted in countries where supply boosts were not included in the programs has not been determined.

4.9 Health impacts measured are those that can be expected to be affected by the program and are measured appropriately.

This final assumption relates to our ability to measure the true impacts of a CCT program and holds if the program actually addresses the factors that affect both health care decisions and health outcomes. For example, CCT programs target the reduction of infant and maternal mortality; however, depending on the context, these outcomes may be more influenced by availability and use of a quality hospital during birth than factors that are more directly affected by a CCT program, such as maternal nutritional status. In addition, the evaluation instruments used to gauge program effectiveness are assumed to be appropriate and sufficient to capture the changes that arise as a result of the intervention. For example, in poor regions where a significant portion of births occur outside health facilities, the fact that CCT programs evaluate infant and maternal mortality based on facility reports, rather than sample surveys, may lead to significant under-estimation of these measures.

Moreover, CCT program evaluations have given very limited attention to the impact on health-related behaviors, attitudes and household decision-making or how these factors contribute or limit impacts on outcomes. As is clear from the evaluation documents, the majority of the program evaluations have focused on measuring the effects on specific

health indicator outcomes, such as incidences of illnesses, child growth and outputs such as utilization rates of public facilities for preventative, curative and prenatal care. Granted, improvements in these outcomes and outputs were listed as goals by most programs, so determining the program's impact is an important first step. However, many of the outputs, especially use of preventive and prenatal care are actually conditions for receipt of transfers, so it should be expected that the programs lead to increases. Moreover, outcomes such as illnesses and child growth are observed more or less at the end of a black box, with evaluations providing little understanding of exactly how the program operates to effect such changes in outcomes.

For example, both Oportunidades and RPS were found to have improved health outcomes among young children, but it is not clear exactly which components were important in achieving the improvements. These impacts could be due to the receipt of nutritional supplements, increased use of preventative and prenatal care services, increased food consumption, increased knowledge in topics covered by the health information lectures (such as proper hygiene and food preparation, best practices for breastfeeding and treatment of diarrhea) or even increased coverage and timeliness of vaccinations through increased supply of preventative care services. Understanding the role of such factors in influencing outcomes is critical for developing more effective programs, yet few evaluations have attempted such an investigation.

A rare and recent exception to this gap in the evaluations is the evidence that has emerged about the limited role of the nutritional supplements in improving child growth and nutrition in Oportunidades. Neufeld et al (2005) studied the proper preparation and consumption of the nutrition supplements by children under five in urban beneficiary households. Their findings suggest that the supplements could have a greater impact if a greater quantity was provided to families (because doses are commonly reduced because sharing with non-targeted household members is common) and that more attention should be paid to ensuring on-time delivery of the supplements to localities and in how the information about proper preparation is delivered to beneficiaries. In addition, the form of some nutrients was found to be less easily absorbed, leading to modifications of the formulation (IDB 2005). Thus, without careful study of the delivery process and use of the supplements by beneficiaries, these issues would continue to limit the effectiveness of at least one component of the program.

The fact that the nutrition supplements were often not prepared and consumed properly by beneficiaries in Mexico suggests that the growth improvements observed among young children were mainly due to increases in the quantity and variety of household food consumption. However, a key component missing in this analysis is an understanding of how well other information presented in the lectures was received and incorporated into beneficiary households. Qualitative studies have indicated that the information transfer in the health lectures was minimal, in part due to the fact that they were at first only delivered in Spanish and not in native languages. However, behavioral changes could also be studied quantitatively through household surveys that collect detailed information about breast and supplemental feeding practices of young children, treatment of incidences of diarrhea and other illnesses (not just documenting occurrences), immunization records for young children, outcomes of current and past

pregnancies (not just prenatal care) and other information related to the content of health lectures. The Demographic and Health Surveys (DHS) can provide a good model from which to base the design of future evaluations surveys as their questions have been tested repeatedly in a variety of social and cultural settings.

Finally, although the program rationale and effect model indicate that the reduction of out of pocket and opportunity costs associated with health care seeking is both the principal mechanism to increase utilization and thus one of the outcome variables on which one would expect to observe impact, no evaluation has analyzed these aspects. Many other examples can be cited.

While the general omission of health impacts in the evaluation framework could be a result of the lack of linkages between data sets for some types of outcomes such as nutritional status or use of specific types of preventive care, in most cases, it seems that the health and nutrition objectives were essentially afterthoughts thought to be “good” but not meriting more in-depth analysis. The Colombia, Mexico and Nicaragua programs modeled the effects that the programs were to have on poverty, inequality, consumption and school attendance in order to provide a framework for assessing the results of the evaluations. However, few programs modeled health effects in a similar way prior to the program design. The exceptions are the Honduras program, which modeled the demand for health services, and Mexico and Nicaragua programs, which modeled nutrition effects. Thus, much could be gained by prioritizing the measurement and evaluation of health impacts from the beginning of program and evaluation design.

V. CONCLUSIONS AND RECOMMENDATIONS

CCT impact evaluations provide unambiguous evidence that financial incentives work to increase utilization of key services by the poor. Further, the evaluations indicate that cash transfers, accompanied by information, social support, weight monitoring and micronutrient supplementation, can stimulate healthier feeding practices and improve young children’s nutritional status dramatically, particularly the incidence of stunting. The Mexican program suggests that adult health may also benefit.

The multi-dimensionality of CCT program benefits is an added attraction. Unlike specific demand-side incentives, such as incentives for HIV testing or compliance with a regimen of medication, CCT programs recognize that the barriers to better health and service use are part of a broader problem of household resource scarcity. Findings suggest that the poorest households must reach a minimum threshold of food consumption before they are able to make other investments in their well-being. Further, better nutritional status increases the effectiveness of health treatments. And because CCT programs are targeted to poor children, the cohort gains associated with preventive care and schooling are irreversible, even if a program must be halted. Moreover, cost information from Latin America indicates that the programs are affordable.

However, the mixed picture with respect to outcomes – vaccination, nutritional status and, where we have data, morbidity and mortality – suggest that encouraging utilization when

services are of poor quality may not produce the expected effects. Moreover, the mixed results suggest that assumptions about needs, household decision-making and causal relationships might not be entirely correct and thus our expectations for impacts, given the current program designs, may be incorrect.

Financial incentives are a blunt instrument that can also have many unintended effects, such as those observed on fertility in Honduras. For this reason, it is important to design incentives carefully. This is particularly important with regard to the health and nutrition components as the evidence reviewed seems to indicate an under-design of these aspects of the CCT programs.

A key design question relates to identifying the marginal benefit of conditioned over unconditioned transfers. Monitoring conditionality is costly and thus it is important to determine whether conditions are necessary and if so, whether enforcement is critical. In some cases, it is probable that only increased income was needed and that poverty was the entire problem, but this should be carefully considered in the design stage. Another important factor to consider in this regard is the baseline status of outcomes. A low baseline means better results; otherwise CCT programs may not have any impact and/or be cost-ineffective as a strategy to increase demand or provider productivity.

A second critical question relates to the relative cost effectiveness of investing on the supply versus the demand-side within the health system. Supply and demand are jointly determined and, while paying poor households to use preventive services works at increasing utilization, it is still unclear what happens at the health post. If quality decreases, or non-beneficiaries are crowded out, the programs may pay too much for the care that beneficiaries receive. In other words, negative spillovers in service quality from demand-side programs may be greater than the net gain to beneficiaries. On the other hand, increased demand can encourage improvements in efficiency and quality.

Regardless, these questions stress the need for an assessment of the supply-side and ex ante modeling of the demand for health care. The issue of cost-effectiveness should also be considered with respect to nutrition; program designers must assess how well a CCT program performs relative to an in-kind transfer or food price subsidy. In addition, the ex ante effects on utilization, consumption and nutrition should be modeled to determine the burden of conditioned services for an average household. Alvarez et al (2006) show that even with the many conditions imposed on households, most households (especially the poorest) perceive a net gain as only the least poor among beneficiaries drop out voluntarily. However, perhaps welfare gains could be increased by reducing burdens with little loss in impacts.

A final word on evaluation. This paper has found limited analysis of the health effects of CCT programs outside of Mexico, and even less of the impact the programs had on health providers themselves. Expanding the scope of future CCT program impact evaluations to include effects on outputs such as knowledge, behaviors, service access and supply can have large benefits as far as improving our understanding of how the program works to achieve improvements. However, the choice to expand questionnaires and/or add instruments most likely involves a trade off of depth in other areas in order to

maintain data quality. It is often the case that multipurpose surveys sacrifice depth for breadth and it would be counterproductive to reduce quality in other areas of evaluation.

However, it is important to note that many of the program evaluation surveys already include questions that could be used to expand the scope of impact evaluations, yet the data have not been used. Perhaps opening access to the data to more researchers will lead to further studies, the results of which could be used to guide the development and design of future evaluations and instruments. For example, the facility surveys and the measures of mothers' knowledge of proper care for diarrhea collected in for the evaluation of FA could provide helpful insight into the underlying health care supply and knowledge that exist for poor households and repeated measures could be used to study the programs impacts on these factors. Moreover, the experience of working with the data could help inform researchers about how such instruments can be improved upon for use in other programs and situations.

Probably one of the largest gaps in the impact evaluations discussed was the lack of information about health care service supply. The quality and quantity of health care available to poor households could have a large effect on the health status not only of children but also of their parents. Improving the tools and strategies used to measure the quality of service delivery was the topic of a recent workshop at the World Bank (April 27-28, 2006) in which participants shared their experiences in collecting and studying a variety of aspects related to health care service provision. These approaches included expenditure tracking, estimating efficiency and quality through administrative records of cases, supplies and equipment, measuring quality via direct observation of consultations and vignettes and capturing demand factors through exit polls. One of the most common observations made by participants was the fact that the utility of facility-level data on supply and quality is increased when matched to household level data on health outcomes. Thus, CCT program impact evaluations appear to be the most appropriate setting in which to use supply-side instruments and develop approaches for studying the relationship between demand and supply factors of health outcomes.

Taking these factors into account during the design of a CCT program provides a unique and important opportunity to improve the effectiveness of these programs and, given the targeting to the poor, the effectiveness of the health system itself on health and nutrition status.

BIBLIOGRAPHY

General

Caldés, Natalia, D.Coady and JA. Maluccio. 2004. The Cost of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America. Discussion Paper 174, Washington, DC: IFPRI.

Coady, David, Margaret Grosh, John Hoddinott.

Handa, Sudhanshu. 2002. Raising Primary School Enrollment in Developing Countries: The Relative Importance of Supply and Demand. *Journal of Development Economics* 69: 103-128.

Handa, Sudhanshu and Benjamin Davis. 2006. Design choices for conditional cash transfers in Latin America and the Caribbean. Forthcoming in *Development Policy Review*.

Heinrich, Carolyn J. 2005. Demand and Supply-Side Determinants of Conditional Cash Transfer Program Effectiveness: Improving the First-Generation Programs. OVE Working Paper. IDB, Washington, D.C.

Inter-American Development Bank. Social Protection for Equity and Growth. Washington, DC: IDB and Johns Hopkins University Press, 2000.

Legovini, Arianna, and F. Regalia. 2001. "Targeted Human Development Programs: Investing in the Next Generation." Inter-American Development Bank, Sustainable Development Department, Poverty and Inequality Advisory Unit. Washington, D.C.

Lustig, Nora, ed. *Shielding the Poor: Social Protection in the Developing World*. Washington, DC: Brookings Institution Press and IDB, 2001.

Rawlings, Laura B. and G.M. Rubio. 2003. *Evaluating the Impact of Conditional Cash Transfer Programs: Lessons from Latin America*. The World Bank: Washington, D.C.

Sedlacek, Guilherme, N. Ilahi and E. Gustafsson-Wright. 2000. "Targeted Conditional Transfers Programs in Latin America: An Early Survey", The World Bank: Washington, D.C.

Strauss, J and D Thomas. 1995. Empirical modeling of human resources. In *The Handbook of Development Economics*, ed. J. Behrman and T.N. Srinivasan. New York: North-Holland.

Oppenheimer, Stephen J. 2001. "Iron and Its Relation to Immunity and Infectious Disease," *Journal of Nutrition*. Vol. 131: pp. 616S-635S.

Brazil:

Belik, Walter and Mauro Del Grossi, 2003, “Brazil’s Zero Hunger Program in the Context of Social Policy” Unicamp.

Bourguignon, François, F. H. G. Ferreira and P. G. Leite, 2002, “Ex-ante Evaluation of Conditional Cash Transfer Programs: The Case of Bolsa Escola” William Davidson Working Paper Number 516. University of Michigan Business School. September 2002.

Bourguignon, François, F. H. G. Ferreira and P. G. Leite, 2003, “Conditional Cash Transfers, Schooling and Child Labor: Micro-Simulating Bolsa Escola”, DELTA, Departement et Laboratoire d’Economie Theorique et Appliquee. Working Paper 2003-07.

Governo Federal do Brasil, 2004, “Principais ações implementadas pelo Programa Fome Zero - Balanço de 2003”.

Ministerio da Educacao and Secretaria do Programa Nacional de Bolsa Escola, 2003, “Conditional cash transfer programs: Operational experiences – O Programa Nacional de Bolsa Escola”, World Bank Workshop, March 12, 2003.

Morris SS, Olinto P, Flores R, Nilson EAF, and Figueiró AC, 2003, “Conditional cash transfers are associated with a small reduction in age-related weight gain in children in Northeast Brazil”. London School of Hygiene & Tropical Medicine, London, UK, International Food Policy Research Institute, Washington DC, USA, Emory University, Atlanta GE, USA, Ministry of Health, Brasília DF, Brazil, Instituto Materno-infantil de Pernambuco, Recife PE, Brazil. October 2003.

Olinto, Pedro, R. Flores, S. Morris, and A. Veiga, “The Impact of the Bolsa Alimentação Program on Food Consumption, Preliminary and incomplete (not for citation)”, International Food Policy Research Institute, Emory University, The London School of Hygiene and Tropical Medicine, Instituto de Estudos do Trabalho e Sociedade. July 2003.

Olinto, Pedro, E. Nilson, S. Morris, R. Flores, A.C. Figuero, and A. Veiga, 2001, “Using non-experimental methods to evaluate a Conditional Cash Transfer program: The case of Bolsa Alimentação”.

The World Bank, 2002, “Technical Note on Projeto Fome Zero”.

Yap, Yoon-Tien, G. Sedlacek and P. Orazem. 2001. “Limiting Child Labor Through Behavior-Based Income Transfers: An Experimental Evaluation of the PETI Program in Rural Brazil”. World Bank, Washington, D.C.

Colombia:

Székely, Miguel, Marianne Hilgert and Amanda Glassman. Criterios para estimar el valor de un subsidio condicionado en Colombia. Technical note, IDB, May 26, 2000.

Székely, Miguel and Marianne Hilgert. Assessment note on poverty, employment, school attendance in Colombia. Technical note, IDB, February 4, 2000.

Unión Temporal IFS-Econometría S.A.-SEI. Evaluación del Impacto del Programa de Familias en Acción: Informe Metodológico. Bogotá: DAPR-FIP y DNP, 2000.

Unión Temporal IFS-Econometría S.A.-SEI. Evaluación del Impacto del Programa de Familias en Acción: Informe de Focalización. Bogotá: DAPR-FIP y DNP, 2001a.

Unión Temporal IFS-Econometría S.A.-SEI. Evaluación del Impacto del Programa de Familias en Acción: Informe Final Línea de Base. Bogotá: DAPR-FIP y DNP, 2001b.

Unión Temporal IFS-Econometría S.A.-SEI. Evaluación del Impacto del Programa de Familias en Acción: Informe del Primer Seguimiento. Bogotá, Colombia: DAPR-FIP y DNP, August 31, 2004.

Honduras:

Alvarez, Carola, A. Santiago and J. Todd. 2005. PRAF: The Family Allowance Program. Draft, IADB.

Flores, Rafael, S. Morris, P. Olinto, J. Medina and O. Neidecker, 2003, "Evaluation of the Family Allowance Program (PRAF) in Honduras: Health and Nutrition Impacts", Emory University, London School of Hygiene and Tropical Medicine and International Food Policy Research Institute.

Garrido Merlano, Victor, 2003, "Evaluación del Sistema de Información para el Registro de Beneficiarios (Borrador)", Informe de Consultoría, Proyecto BID 1026 / SF-HO PRAF Fase II – Honduras.

IFPRI, 1999, "First Report: Activities completed up to July 31st, 1999", IFPRI.

IFPRI, 2000. "Segundo Informe: Propuesta de implementación del Proyecto PRAF/BID - Fase II", Unidad Coordinadora del Proyecto (UCP), IFPRI.

IFPRI, 2000. "Second Report: Implementation Proposal for the PRAF/ IDB Project – Phase II", Project Coordinating Unit and IFPRI, March 27, 2000.

IFPRI, 2000, "Tercer Informe: Monitoreo y Evaluación del Sistema", Junio 15, 2000.

IFPRI, 2000, "Third Report: Monitoring and Evaluation System", June 15, 2000.

IFPRI, 2001, "Cuarto Informe: Proyecto PRAF/BID Fase II: Análisis de la Situación Encontrada luego del Arranque del Proyecto" IFPRI.

IFPRI, 2001, "Fourth Report of the PRAF Series PRAF/ IDB Phase II: Analysis of the Situation Before the Beginning of Distribution of Vouchers and Project Implementation", IFPRI, Washington, D.C. September 2001.

IFPRI, 2003, “Sexto Informe: Proyecto PRAF/BID Fase II: Impacto Intermedio” IFPRI.

Leiva, Miriam, 2001, “General Background Information on PRAF II Project”, Proyecto Contrato de Préstamo BID 1026 SF-HO – Praf Fase II. Second External Advisory Committee Workshop. Honduras, July 18-19, 2001. IFPRI Website.

Morris, Saul, 2001, “Baseline Health Results”, Second External Advisory Committee Workshop. Honduras, July 18-19, 2001. IFPRI Website.

Morris, Saul, and R. Flores, 2002, “School Height Censuses Are Reliable and Valid Tools for Small-Area Targeting of Nutrition Interventions in Honduras”, The American Society for Nutritional Sciences J. Nutr. 132:1188-1193.

Morris, Saul S, R. Flores, P. Olinto, and J.M. Medina, 2004, “The Family Allowance Program, Honduras: Interim Impact Evaluation”. IFS, January 2004.

Morris, Saul S, R Flores, P Olinto and JM Medina, 2004, Monetary incentives in primary health care and effects on use and coverage of preventive health care interventions in rural Honduras: cluster randomised trial. The Lancet 2004; 364:2030-2037.

Jamaica:

Bronte-Tinkew, Jacinta, 1999, “Family Structure, Resource Dilution and Child Nutrition: an Analysis of Jamaica”, Population Research Institute - The Pennsylvania State University, Working Paper 99-05. March 1999.

Fawcett, Caroline, 2004, “Social Safety Net Report”, IMPAQ International, LLC, Mathematica Policy Research and the Government of Jamaica, April 2004.

Government of Jamaica, “National Poverty Eradication Programme, <http://www.npep.org.jm/index.html>”, http://www.npep.org.jm/Projects_Programmes/Ministry_of_Labour_and_Social/P_ATH/p_ath.html.

<http://www.npep.org.jm/Papers/PATH/path.html>.

IADB, 2001, “Jamaica. Proposal for a loan and nonreimbursable technical-cooperation. funding for a social safety net reform program”, PR-2612 - 11 September 2001.

Mathematica. Targeting Report and Evaluation Proposal [PENDING FROM WB].

World Bank, 2003, “Programme of Advancement through Health and Education (PATH)”, March 12.

Mexico:

Adato, Michelle. 2000. "The Impact of PROGRESA on Community Social Relationships." September. International Food Policy Research Institute, Washington, D.C.

Adato, Michelle, B. de la Brière, D. Mindek, and A. Quisumbing. 2000. "Final Report: The Impact of PROGRESA on Women's Status and Intrahousehold Relations." July. International Food Policy Research Institute, Washington, D.C.

Adato, Michelle, David Coady and Marie Ruel. 2000. "An Operations Evaluation of PROGRESA from the Perspective of Beneficiaries, Promotoras, School Directors, and Health Staff". International Food Policy Research Institute, Washington, D.C.

Álvarez, Carola, Florencia Devoto and Paul Winters. 2006 "[Why Do the Poor Leave the Safety Net in Mexico? A Study Of the Effects of Conditionality On Dropouts](#)" working paper No. 2006-10, American University, Washington, D.C.

Attanasio, Orazio, C. Meghir, and A. Santiago, 2001, "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA", EdePo – IFS – November 2001.

Attanasio, Orazio, C. Meghir, and M. Szekely, 2003, "Using Randomized Experiments and Structural Models for "Scaling Up": Evidence from the PROGRESA Evaluation". EdePo – IFS – May 2003.

Bautista, Sergio, S. Martínez, S. Bertozzi, and P. Gertle, 2003, "Evaluación del efecto de Oportunidades sobre la utilización de servicios de salud en el medio rural - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades", Instituto Nacional de Salud Publica (INSP), marzo de 2003.

Behrman, Jere R. and Petra E. Todd. 1999a. "A Report on the Sample Sizes Used for the Evaluation of The Education, Health and Nutrition Program (PROGRESA) of Mexico." International Food Policy Research Institute. Washington, D.C.

Behrman, Jere R. and Petra E. Todd. 1999b. "Randomness in the Experimental Samples of PROGRESA (Education, Health and Nutrition Program)." International Food Policy Research Institute. Washington, D.C.

Behrman, Jere and John Hoddinott. 2000. "An Evaluation of the Impact of PROGRESA on Pre-school Child Height". International Food Policy Research Institute. Washington, D.C.

Behrman, Jere, P. Sengupta and P. Todd. 2000. "The Impact of PROGRESA on Achievement Test Scores in the First Year," International Food Policy Research Institute. Washington, D.C.

Behrman, Jere, P. Sengupta, and P. Todd. 2001. "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment." April. University of Pennsylvania and the International Food Policy Research Institute, Washington, D.C.

Coady, David. 2000. "The Application of Social Cost-benefit Analysis to the Evaluation of PROGRESA." December. International Food Policy Research Institute, Washington, D.C.

Coady, David and R. Harris. 2000. "A General Equilibrium Analysis of the Welfare Impact of PROGRESA." April. International Food Policy Research Institute, Washington, D.C.

Coady, David and S. Parker. 2002. A Cost-Effectiveness Analysis of Demand- and Supply-Side Education Interventions: The Case of ProgresA in Mexico. March. IFPRI, Washington, D.C.

Davis, Benjamin, S. Handa, M. Ruiz Arranz, M. Stampini, and Paul Winters, 2002, "Conditionality and the impact of program design on household welfare: Comparing two diverse cash transfer programs in rural Mexico", IDB, FAO, Harvard University, and Scuola Sant'Anna, Pisa, October, 2002.

Escobar, Agustín and M. González de la Rocha, 2002, "Evaluación Cualitativa del Programa de Desarrollo Humano Oportunidades, Seguimiento de impacto 2001 – 2002, Comunidades de 2,500 a 50,000 habitantes - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades" CIESAS, diciembre de 2002.

Escobar Latapí, Agustín y Mercedes González de la Rocha (2005) "Evaluación cualitativa del Programa Oportunidades en zonas urbanas, 2003" en Hernández Prado, Bernardo y Mauricio Hernández Ávila (ed.) *Evaluación externa de impacto del Programa Oportunidades 2003*. INSP – CIESAS, 2005.

Gertler, Paul J. 2000. "Final Report: The Impact of PROGRESA on Health". International Food Policy Research Institute. Washington, D.C.

Gertler, Paul J. and Simone Boyce, 2001, An experiment in incentive-based welfare: the impact of ProgresA on Health in Mexico. Processed.

Gertler, Paul J. and L.C. Fernald. 2004. The Medium Term Impact of Oportunidades on Child Development in Rural Areas. Processed.

Gutiérrez, Juan Pablo, S. Bertozzi, and P. Gertler, 2003, "Evaluación de la identificación de familias beneficiarias en el medio urbano - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades", INSP, marzo de 2003.

Gutiérrez, Juan Pablo et al (2005) "Impacto de Oportunidades en la morbilidad y el estado de salud de la población beneficiaria y en la utilización de los servicios de salud. Resultados de corto plazo en zonas urbanas y de mediano plazo en zonas rurales" en

Hernández Prado, Bernardo y Mauricio Hernández Ávila (ed.) *Evaluación externa de impacto del Programa Oportunidades 2004*. INSP – CIESAS, 2005.

Handa, S., M-C. Huerta, R. Perez, and B. Straffon. 2000. “Final Report: Poverty, Inequality, and "Spill-Over" in Mexico's Education, Health, and Nutrition Program.” April. International Food Policy Research Institute, Washington, D.C.

Hernández, Bernardo, J.E. Urquieta, F. Meneses, M.C. Baltazar, and M. Hernández, 2003, “Evaluación del cumplimiento de metas, costos unitarios y apego del Programa Oportunidades a las Reglas de Operación - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades”, INSP, Marzo de 2003.

Hernández, Bernardo, J.E. Urquieta, F. Meneses, M.C. Baltazar, and M. Hernández, 2003, “Anexo - Evaluación del cumplimiento de metas, costos unitarios y apego del Programa Oportunidades a las Reglas de Operación - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades”, INSP, Marzo de 2003.

Hernandez, Gonzalo, G. Merino, A. Santiago, and M. Szekely, 2001, “Design and Evaluation of Social Programmes - Mexico”, Ministry of Social Development and SEDESOL.

Hoddinott, J., E. Skoufias, and R. Washburn. 2000. “The Impact of PROGRESA on Consumption: A Final Report. September.” International Food Policy Research Institute, Washington, D.C.

Instituto Nacional de Salud Pública and CIDE, 2002, “Introducción. Evaluación de Resultados de Impacto del Programa de Desarrollo Humano de Oportunidades”, Septiembre 2002.

INSP and CIESAS, 2003, “Introducción. Evaluación de Resultados de Impacto del Programa de Desarrollo Humano de Oportunidades”, Marzo 2003.

Meneses G., Fernando et al (2005) “Evaluación del cumplimiento de las metas, los costos unitarios y el apego del Programa a las Reglas de Operación 2004” en Hernández Prado, Bernardo y Mauricio Hernández Ávila (ed.) *Evaluación externa del impacto del Programa Oportunidades 2004, Aspectos económicos y sociales*. INSP – CIESAS, 2005.

Parker, S., and E. Skoufias. 2000. “Final Report: The Impact of PROGRESA on Work, Leisure, and Time Allocation.” October. International Food Policy Research Institute, Washington, D.C.

Rubalcava Peñafiel, Luis N. and G. Teruel Belismelis, 2003, “Análisis sobre el cambio en variables demográficas y económicas de los hogares beneficiarios del Programa Oportunidades 1997-2002 - Evaluación de Resultados de Impacto del Programa de Desarrollo Humano Oportunidades”, INSP and CIDE, marzo de 2003.

Sadoulet, Elisabeth, F. Finan, and A. de Janvry, 2002, "Decomposing the Channels of Influence of Conditional Transfers in a Structural Model of Educational Choice", University of California at Berkeley, December 2002.

Schultz, T. Paul. 2000a. "Impact of Progresa on School Attendance Rates in the Sampled Population" International Food Policy Research Institute: Washington, D.C.

Schultz, T. Paul. 2000b. "School Subsidies for the Poor: Evaluating a Mexican Strategy for Reducing Poverty." June. International Food Policy Research Institute, Washington, D.C.

Schultz, T. Paul. 2000c. "Final Report: The Impact of PROGRESA on School Enrollments." April. International Food Policy Research Institute, Washington, D.C.

Schultz, T. Paul. 2001. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program." August. Yale University Economic Growth Center, New Haven, CT.

Skoufias, E. 2001. "PROGRESA and its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI." December. International Food Policy Research Institute, Washington, D.C.

Skoufias, E., B. Davis, and J. Behrman. 1999. "Final Report: An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico." International Food Policy Research Institute: Washington, D.C.

Skoufias, E., B. Davis, and S. de la Vega. 1999. "An Addendum to the Final Report: An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico. Targeting the Poor in Mexico: Evaluation of the Selection of Beneficiary Households into PROGRESA." December. International Food Policy Research Institute, Washington, D.C.

Skoufias, Emmanuel, and B. McClafferty. 2001. "Is Progresa Working? Summary of the Results of an Evaluation by IFPRI" International Food Policy Research Institute: Washington, D.C.

Teruel, G., and B. Davis. 2000. "Final Report: An Evaluation of the Impact of PROGRESA Cash Payments on Private Inter-Household Transfers." August. International Food Policy Research Institute, Washington, D.C.

Nicaragua:

Adato, Michelle and T. Roopnaraine. April 2004. Informe Preliminar. Sistema de Evaluación de la Red de Protección Social de Nicaragua: Un Análisis Social De La Red De Protección Social. IFPRI, Washington, D.C.

IFPRI. 2001a. "Informe: Sistema de Evaluación de la Fase Piloto de la Red de Protección Social de Nicaragua." processed.

-----, 2001b. "Sistema de Evaluación de la Fase Piloto de la Red de Protección Social de Nicaragua: Línea de Base 2000." processed.

-----, 2002a. "Informe Final - Sistema de Evaluación de la Fase Piloto de la Red de Protección Social de Nicaragua: Evaluación de Impacto." processed.

-----, 2002b. "Informe Final - Sistema de Evaluación de la Fase Piloto de la Red de Protección Social de Nicaragua: Evaluación de Focalización." processed.

Annex A

Evaluation design and its limits

The focus of this section is to discuss the impact evaluations conducted for each of the programs, highlighting both the strategies employed in each evaluation to identify the program impacts, but also the limitations of the chosen designs, the instruments used to measure outcomes and the scope in relation to health outcomes.

Impact Evaluation Designs. Theoretically, impact evaluations are quite simple; outcomes of beneficiaries are compared to the outcome that would have resulted in the absence of the program, and the difference between the two is the impact of the program. However, in practice conducting such a simple comparison is virtually impossible as the counterfactual does not exist for beneficiaries due to the simple fact that they were beneficiaries. Instead, a control or comparison group must be constructed, which can be reasonably expected to represent the condition of the beneficiary group in the absence of the program. In clinical trials of new medicines, the experimental design in which eligible individuals are randomly allocation of to treatment and control groups is considered the gold standard for constructing a counterfactual to compare to the treated group. Anything less than individual random allocation has the potential to create selection bias in estimated impacts, especially when beneficiaries can choose whether or not to participate, or when beneficiaries are selected because they possess a particular set of characteristics. Any time there are differences in observed or unobserved characteristics between beneficiaries and non-beneficiaries, it is necessary to account for these differences in the estimation of impacts, which at times doing so can be nontrivial.

Given the nature of the interventions in CCT programs (cash transfers to households and in some cases increased funding and resources to health and education service providers), individual randomization is not feasible. Even random allocation at the household level poses challenges. Concerns for equity and minimizing social unrest make it infeasible to include and exclude households randomly within a small community. Moreover, when the program includes supply-side interventions, although it is possible to randomly allocate resources across providers, it is impossible to exclude a portion of households within a provider's catchment area from the improved services. Thus, in order to conduct an evaluation that will provide reasonable measures of the effects of the program, alternative strategies must be employed to identify an appropriate counterfactual for beneficiary households.

Another important factor that needs to be considered when conducting evaluations is the issue of representation, both of the population and of the program, in the sample used for evaluation. In order for estimated impacts to be valid for the full population of beneficiaries, the evaluation sample used to measure the impacts of the program must be representative of the population of beneficiaries. If the sample is not representative, the results of the impact evaluation are only valid for the subgroup represented by the sample and cannot be assumed to hold for the entire beneficiary population (Stock and Watson 2003). Moreover, when an evaluation is conducted on a pilot version of the program, the impacts cannot be assumed to be externally valid and hold for an expanded version if the pilot phase is different from the expanded program. It is often the case that pilot phases

of programs identify problems with design and operations and the information gained from the pilot phase is used to improve the program through modification prior to an expansion.

Each program discussed in this paper dealt with the key problem, constructing a counterfactual for evaluation, in slightly different ways and few were able to ensure that the results were valid estimates of the impact of the program for the full population of beneficiaries or that they could be reasonably expected to occur in an expanded version of the program. The impact evaluations of OPORTUNIDADES, PRAF and RPS relied on the random allocation of the program at the community level from a pre-selected set of communities that were determined to be eligible for the intervention.²⁸ FA and PATH utilized quasi-experimental designs, employing matching techniques to construct a counterfactual group from a pool of communities (in the case of FA) or households (in the case of FA) excluded from the program(s). In addition, OPORTUNIDADES used quasi-experimental techniques for estimating impacts after 6 years of program operation (after the control group had been incorporated into the program as well). However, as will be elaborated below, each of the evaluations were conducted on pilot phases and were not conducted on representative samples and therefore the portability of results could be limited.

Randomization at the community level, although not the gold standard for evaluation can result in the construction of two groups with similar distributions of observable characteristics where the group that would not receive benefits from the program can serve as a control for the group that did. With this design, the differences-in-differences estimator can be used with panel data to account for initial differences in observable characteristics between the two groups and can produce unbiased estimates of the effect of the program as long as it can be assumed that both groups faced similar macroeconomic environments, other than the program (Stock and Watson 2003). A significant advantage to random assignment at the community level is that it avoids selection bias that can result when non-eligible or non-participant households are used as counterfactuals for beneficiaries as in quasi-experimental designs. However, there is always the possibility that differences in unobservable community characteristics and dynamics could have a large influence over the way in which the program operates and the impact that it has on its beneficiaries. When unobservable characteristics are correlated with observables, this potential problem can be mitigated, but it is always an important factor to keep in mind.

Quasi-experimental designs are often implemented when randomization (even for a brief period of one to two years) is not feasible, which could occur for many reasons. One possible reason for not randomizing is that governments may not want to randomly exclude otherwise eligible individuals due to concerns about equity. Another possible

²⁸ The term ‘community’ is used rather loosely here to refer to the smallest unit (above the household) used for allocating the program. In the case of PRAF, the program was allocated randomly to municipalities (which contain multiple communities) due to the fact that health and education service providers had catchment areas which spanned municipalities. In RPS, the program was randomized across a set of *comarcas*, census units. Only in the case of OPORTUNIDADES was the randomization actually at the community (locality) level.

reason is that countries may lack sufficient prior information at the community level that would allow for similar control and treatment groups to be constructed. Quasi-experimental designs also arise if the evaluation was not initially included in the original program design and was implemented after the program began, or simply was delayed due to administrative, financial or political issues. Thus, the quasi-experimental design can be a good compromise between conflicting interests or can provide the means to conduct an evaluation even when it was not included in the original design. The main disadvantage to this design is that it is nearly impossible to completely eliminate all forms of selection bias in the control and treatment groups, especially if the program is national, although if done carefully, such biases can be drastically reduced.

A. OPORTUNIDADES (MX)

Evaluations of OPORTUNIDADES were conducted using both administrative data for public health facilities, qualitative studies at the facility and community level as well as experimental panel household data collected through household surveys in a sample of rural beneficiary communities and quasi-experimental panel household data collected in urban areas. The administrative data was mainly used to study changes in utilization rates of public clinics for both preventative and curative care, while the qualitative studies were able to capture provider and beneficiary perceptions about the program and its components as well as changes that occurred due to the program.

The rural panel household survey data was used for the first of the impact evaluations due to the fact that the program began in rural areas and covered a set of 506 communities located in the central part of the country were selected for the evaluation. Given the fact that budget and administrative constraints resulted in the need for incorporation to be phased in over time across the country, communities within the evaluation sample could be randomly allocated into two groups, one that would receive program benefits immediately and other group that would begin to benefit from the program after 1 ½ to 2 years after the treatment group and could serve as a control for the earlier treated group (Berhman and Todd 1999). Panel household data for the evaluation was collected through repeated household surveys, beginning with a baseline observation in 1997. Follow-up survey rounds were done in October 1998, May 1999, November 1999, May 2000, November 2000 and most recently in 2003. Unlike most other program evaluations, the surveys were administered to all households in each community (a census was conducted in each survey round) instead of on a sample of households.

The control group began to benefit from the program in the fall of 1999, so the experimental data spans the period between the baseline survey in 1997 and the November 1999 survey. Initial evaluations were conducted using the difference-in-differences estimator after 1 ½ to 2 years of program operation with the experimental data. Due to the location, the communities in the evaluation sample were not necessarily representative of the entire beneficiary population, however, they had the

advantage that were more easily accessible for repeated surveying, which helped to reduce costs associated with the evaluation.²⁹

Later evaluations were conducted after six years of operation in rural areas and one year of operation in urban areas that utilized a quasi-experimental design in which a control group was constructed through propensity score matching. In rural areas, a set of rural communities was selected to serve as a comparison group based on similarities to the treated communities and households were then matched through propensity scores. In the urban areas, where incorporation was conducted much differently, the comparison group was also constructed through carefully selection of non-beneficiary households and propensity score matching. To date, the urban evaluation sample covers the period 2002-2003.

B. RED DE PROTECCIÓN SOCIAL (NI)

The evaluation of RPS in Nicaragua was conducted on the pilot phase of the program and relied on random allocation into the program at the community level. The selection of communities for the pilot involved multiple steps. First, the list of potential communities was narrowed down to those within two administrative departments that were selected due the fact that they had adequate health and education infrastructure and had higher poverty rates higher than the national average, among other criteria. Secondly, marginality indices were constructed for each community in the pre-selected departments using the most recent census data, with the 42 most marginalized communities selected for the pilot phase. A lottery determined assignment of half of the communities, grouped into strata by the marginality index, into the treatment group and half into the control group (Maluccio and Flores 2004). Panel data was collected through repeated household surveys for a random sample of households in both groups, both before (baseline), and at one and two years after, the program began. Thus, the experimental data covers the period fall 2000 to fall 2002.

In 2003, the treated localities were lost the benefits of the conditional cash transfers (but not the supply-side benefits) while the control group began to receive cash transfers and supply-side benefits. To be able to continue the evaluation, another control group was selected and surveyed in 2002, prior to the beginning of the change in interventions so that baseline data for the group could be collected at the same time, and using the same instrument, as the original treatment and control groups. Given the non-experimental selection of the new control group, the evaluation conducted using these data used propensity-score matching.

C. PRAF (HO)

In PRAF, the selection of the communities to participate in the pilot phase involved identifying the municipalities in the country with the lowest average measure of child growth. All municipalities in the country were first ranked according to the height status of first graders through the use of a 1997 census of first graders that collected height and weight

²⁹ In fact, no one has yet determined how well the evaluation sample represents the population of beneficiaries, but it is well known that the central region of the country is less poor than the south and more poor than the north, on average (CITE).

measurements of all children enrolled in first grade. The 70 municipalities with the lowest average height-for-age (haz) z-scores, indicating the poorest growth and development of children, were selected for the pilot phase of the program and for the evaluation.³⁰ The set of 70 municipalities was then organized into four strata according to the average haz and a lottery determined which municipalities from each stratum would be assigned to one of four different treatment protocol (IFPRI 2000a).

The initial program design included four different groups; demand only - a group that would receive household cash transfers only, demand and supply - a group that would benefit only from increased resource flows to health and education service providers, supply only - a group that would receive both cash transfers and have access to health and education facilities with increased resources and a fourth group that would serve as a control group. The idea was that outcomes in households in each of the three different treatment groups could be compared to the control group and to each other so that the impacts from separate components could be investigated. However, administrative difficulties limited the timely implementation of the transfers to health and education providers, which meant that the supply only group received little to no benefit by the time of the follow-up round of data and the demand and supply group only received cash transfers. In addition, PRAF also suffered from large delays in the payment of transfers to households, which further reduced the intensity of treatment in the groups receiving the cash transfers (IFPRI 2003a).

D. FAMILIAS EN ACCIÓN (CO)

FA was initially implemented nationally in 631 municipalities that met four criteria that were used to help to focus the program in needy areas and ensured that adequate infrastructure was available for the program. Specifically, in order to benefit from the program, the communities had to be classified as poor according to an unmet basic needs index constructed on the 1993 census and had a population less than 100,000³¹, have sufficient health and education service infrastructure and at least one bank and have local leaders which supported the program operations and were willing to fulfill any obligations imposed on them by the program (Unión Temporal IFS 2000).

A random sample of beneficiary communities, and then households within selected communities, was taken for the purposes of evaluation. The control group for the evaluation was selected from the 269 municipalities that were initially selected to participate in the program but did not meet at least one of the eligibility criteria and thus were excluded. Propensity score matching was used to identify the municipalities that were most similar to the beneficiary communities, thus minimizing observable differences between the two groups prior to the program intervention (Unión Temporal IFS 2000). Ultimately, the evaluation surveyed a random sample of households from a sample of 122 municipalities, composed of 57 beneficiary and 65 control municipalities. The fact that the municipalities that did not meet at least one of the eligibility

³⁰ In fact, three of the municipalities with the lowest average height for age z-score of first graders had to be excluded from the sample due to political instability. Therefore, the 71-73 ranked municipalities were included as their replacements (IFPRI 2000a).

³¹ In addition, the municipalities could not be the capital of the department or be located in a coffee region that was benefiting from other welfare programs.

requirements were delegated to the control group poses significant challenges for determining the impact of the program, especially if municipalities in the control group lacked access to health services that were equivalent to those in the treated group prior to the intervention.

In addition, the evaluation of FA faced other significant challenges. The baseline evaluation survey was conducted after some municipalities had already begun to receive transfers and most had been informed about the program and their imminent beneficiary status. Thus, “pre-program” measures were likely contaminated as households already had begun to respond to the program conditions, which would tend bias downward the estimated impact. To attempt to overcome these issues, in addition to using propensity-score matching to narrow the control group to those that were similar to the treated group [although, you will recall, they were not similar enough to be beneficiaries] retrospective information about outcomes was collected in the baseline survey to attempt to get more pure measures of pre-program conditions. For some outcomes, retrospective information can help identify any pre-program differences, however, outcomes such as growth and nutrition are basically immeasurable retrospectively and other outcomes such as consumption and health can be strongly biased downward when recalled.

A final note about the FA evaluation is that the program also experienced delays the transfer of payments to households, and so the impacts observed during the evaluation period are not necessarily those that would occur in a more mature program in which payments are delivered regularly on time.

E. PATH (JA)

The evaluation is yet to be completed (due in late June 2006), but also employed a quasi-experimental methodology that utilized propensity score matching to identify a set of comparison group of households that was similar to a set of beneficiary households prior to the program intervention. While community differences such as in FA are not likely to be a large issue (since selection into the program is based only on household characteristics), the evaluation used households that applied for the program but were found to be ineligible to serve as the comparison group. While this approach overcomes issues of differences in intention to participate in the program (all households in the evaluation applied for the program), the fact that the controls were deemed ineligible while the treated households were eligible means that the two sets of households are different from each other. However, eligibility for the program is based on a score (similar to that for OPORTUNIDADES and therefore the cutoff point may be fairly arbitrary, especially at or near the cutoff and therefore households just above or just below the eligibility threshold may not be all that different from each other.

PATH’s impact evaluation was also conducted using quasi-experimental techniques, but is unique from the other four program evaluations in many ways. The evaluation compared outcomes beneficiaries that were at or near the cutoff for eligibility to matched comparison households by drawing from household program application data. The comparison group was strictly composed of ineligible households, and therefore seems unlikely to serve as an appropriate counterfactual, but by restricting the sample of treated

households to those just below the cutoff to be eligible the differences between the two groups are quite minimal (Levy and Ohls 2003). There is the additional advantage with by selecting comparison households from those that applied for the program, the households are likely to be similar across unobservable characteristics, such as an interest in participating in the program. Thus, the evaluation is conducted for the subset of the beneficiary population that was just poor enough to be eligible and can cannot therefore; indicate the magnitude and/or direction of impacts for beneficiaries with much lower eligibility scores (much poorer households). However, the evaluators highlight that the results do demonstrate the marginal impact of the program and what impacts should be expected if the program is expanded by increasing the eligibility through raising the eligibility score cutoff (Levy and Ohls 2003).

Another difference between the PATH and the other programs is that beneficiaries in the other programs first had to reside in eligible communities, but in PATH eligibility was determined strictly from household characteristics. Thus, it is more likely to find beneficiary and non-beneficiary households in similar (if not the same) communities, reducing the potential for program impacts to be confused with the influence of other factors that differ between the two groups.

F. QUESTIONNAIRES

In general, the survey instruments used in evaluations have not included questions that are capable of capturing these factors, and those surveys that have attempted to do so are not nearly as complete to allow for much in-depth analysis of these factors and the relationship to program interventions and outcomes. The early OPORTUNIDADES surveys (1998-2000) were especially lacking in that they only included very minimal questions about the status of and behaviors related to child health. The 2003 rural and urban surveys do include additional instruments to measure factors such as the development of mental and motor skills of young children, diet habits, health status and social activities of adolescents and reproductive history, attitudes and behaviors of women. However, because these modules/instruments only appeared in the latest rural round, the baseline situation is unknown and the estimation of the program impact on these outcomes/outputs could not be done using the early experimental data and instead relied on quasi-experimental techniques.

The RPS questionnaires, which were modified from the LSMS surveys, were a bit more in-depth about child health and illnesses. The child health module included more detailed questions about recent incidences and treatments of diarrhea among young children with many questions focused on whether or not (and why not) treatment was sought and if so, at what type of facility, and the travel and out of pocket costs associated with treatment. The survey did include a question about breastfeeding, but the question was ill posed in that it only asked about the duration of exclusive breastfeeding and not the total duration. Thus, even though the survey collected information about the types of supplemental feeding practices, the lack of complete information about breastfeeding made it difficult to studying the impact of the program on household feeding behaviors related to child nutrition. The RPS child health survey module did record whether or not the child had been weighed at the last well child exam, had received vitamin A and iron supplements

recently and had received critical vaccinations for diseases such as TB, DPT and polio, although dates of vaccinations were not recorded.

The PRAF and FA child health questionnaires collected the richest information about health-related behaviors. This included whether or not the child washed his/her hands before eating (PRAF), duration of breastfeeding, the types of foods provided to the child (especially detailed in FA), if the child was taken to a health center in the past month and the reason for the visit (even if more than 30 days ago), the specific dates on which each vaccination was received, and the specific symptoms of the last episode of diarrhea. In the case of PRAF, the module was only applied to children under the age of three and little program impact was observed for many outcomes due to the lack of timely implementation of many of the program components. These factors limited the use of these questions for studying impacts. In the case of FA, the child health module was applied to all children under the age of 7 and also included additional information about each child's birth weight and height, gestational age at birth, past growth measurements if the child was enrolled in a growth monitoring program and incidences of cough, fevers and other illnesses. While some of this data was studied in the impact evaluation (birth weight and illnesses), most were not studied (or at least reported officially in the evaluation).

Questions related to fertility and maternal health were limited in early rural surveys for OPORTUNIDADES and in the RPS, PRAF and FA questionnaires centered on the number of prenatal check-ups, types of screenings that occurred during exams, and the presence of health care providers at the birth. In the RPS survey, there were many questions related to costs associated with prenatal care, but no information about actual birth outcomes such as the gestational age at birth, birth weights or infant mortality (except in the 2002 round, for the most recent birth only). The sample size was relatively small for some outcomes, so perhaps the exclusion of these types of questions was practical. In the PRAF survey, the fertility and maternal health module was more elaborate, capturing the duration of the last pregnancy and the ultimate outcome, but not birth weight. In the FA questionnaire, the maternal health module was only applied to beneficiary women or heads and spouses of heads and it is unclear if all women of reproductive age were covered.

PRAF, RPS and FA did not collect complete birth histories for women of reproductive age. Instead, such information had to be reconstructed from the household roster and therefore missed any children which were not still alive or living with their mother. Contraceptive use of women was collected in FA (but not reported in the evaluation report) but was not included in PRAF and RPS (with the exception of the 2002 RPS survey, but then only asked of women that had even been pregnant). The rural 2003 survey for OPORTUNIDADES included a comprehensive fertility module that included knowledge and use of contraceptives but did not administer the survey to all women of reproductive age (it was noted that daughters were most likely to be missing). Stecklov et al (2006) found that fertility increased in treated communities in PRAF, but not in RPS and OPORTUNIDADES, and attributed the increase to the fact that nutrition transfers were provided per young child, suggesting that households were reacting to the incentive to have more young children in the household. However, given the short period covered by the surveys and the fact that there was limited information about the proximate determinates of fertility, the source or reason for the impact on fertility it is still unclear. Thus, it seems imperative that future

evaluation surveys for programs which have the potential to affect demographic changes collect as much information as possible that is related to reproductive health and related decision-making within households.

A unique component of the FA maternal module were questions which asked the mother about the proper treatment for diarrhea (specifically, whether food and liquids given to children should increase when the child suffers from diarrhea) and participation in talks which covered a variety of health and nutrition topics. However, this information was not used in the published impact evaluation. These questions are similar to the vignette approach used to evaluate provider competence in studies of health service quality (Das et al 2004, among others) and the use of such instruments could be explored for future evaluations. This point is discussed further below.

Annex B

Evaluation results

Latin American CCT programs include impact evaluation components, a situation that is unprecedented in the region and is primarily due to four factors. As a first implementer, Mexico's Progresa program carried out a state of the art experimental impact evaluation. This led to extensive documentation and dissemination of program results which in turn led to program durability during a period of political transition and generated demand for similar evaluations in countries that followed the Mexican model. Second, as the use of cash transfers as a safety net was relatively novel in most countries, policy-makers found that it was necessary to demonstrate effectiveness in achieving development outcomes *vis a vis* alternative modalities, such as in kind subsidies. In some cases, evaluation results were used as technical justification for the elimination of these subsidies, which suffered from problems of targeting and effectiveness. Third, the rapid expansion of the programs and the magnitude of the beneficiary populations created a need for the evaluations as an accountability tool for government; the use of independent evaluators has protected the programs from charges of politicization.³² Finally, with the exception on Mexico which financed the first phase of Progresa independently, the participation of the Inter American Development Bank and the World Bank appears to have encouraged the inclusion of an impact evaluation, making the experience particularly interesting to examine in light of low evaluation rates of other programs.

Selected Evaluation Results. This section summarizes published information on the health and nutrition-related outputs and outcomes currently available in the literature. As noted in the attached annex, each evaluation's sample design determined the type of analysis that could be conducted on the data and each questionnaire determined what results were measured and reported. In general, difference-in-differences estimates of impacts are reported unless otherwise noted.

Outputs: Utilization. As reviewed earlier, use of preventive health care is thought to be an important input for overall better child health, and such facility use is often strongly correlated with family background indicators such as parental educational attainment. Thus a major expected output of the CCT programs is to mitigate the advantage of socioeconomic background in determining the probability of use of services. All of the programs stipulated utilization of preventive health services as a condition for the receipt of transfers, and as such, these indicators were measured using both administrative and household surveys by all programs.

In the short-term the CCT programs are expected to increase preventive and curative health service utilization. Given the incentive to use public health care providers, a decrease in private health care service utilization might be expected. In addition, if access to preventative health care services improves and service provision is properly carried out, a reduction in the number and gravity of health episodes should be expected which, in turn, should reduce the number of days of hospitalization.

³² Although there is evidence of pre-electoral expansions of the beneficiary populations – though always applying the same targeting criteria -- in Colombia and Mexico.

Gertler and Boyce (2001) investigate the impact of Progresa on visits to public clinics using administrative data, finding that after the introduction of the program in 1997, visit rates to clinics in Progresa localities are on average higher than in non-Progresa localities and the difference grows over time as more Progresa localities begin to provide benefits. About 2.09 more visits per day or about 18.2 percent more visits to clinics are recorded in Progresa areas versus non-Progresa areas (likely made up of both Progresa beneficiaries and spillover effects on non-Progresa beneficiaries – see Handa and Huerta 2001 below). In addition, using evaluation survey data, Gertler and Boyce (2001) suggest that Progresa increased utilization in public clinics by 53 percent overall.

As expected, utilization of conditioned preventive health services increased significantly on average among the poor as a result of CCT programs. The extent of this increase varied in magnitude, generally larger in low use baseline settings such as rural areas and among the poorest households. There are some surprising findings regarding the age distribution of the utilization as well, with effects more pronounced in some programs outside of the youngest age group (0-3 years of age), attributed by analysts to pre-existing high rates of use among this population or reduction in need for curative care due to the transfers' positive impact on nutritional status of mothers (see Table 5).

Although utilization data for children in the Mexican program shows little effect, the use of general preventive care by households appears to have increased dramatically. Overall, Oportunidades increased the average number of preventive health care visits by members of beneficiary families by 20 percent. Qualitative research and administrative data confirms these patterns; Meneses (2005) reports that the main reason for the last visit to the clinic was the Oportunidades-conditioned appointment or health education talks (74%-86%), while illness only represented 17% to 22%.

The greater use of preventive care may also reflect increases in health knowledge and practices, but it is impossible to distinguish the effects of the conditionality from those related to greater knowledge levels. INSP (2003) used monthly clinic-level data from the administrative records of public clinics operated by IMSS-Solidaridad finding that the proportion of preventive consultations increased at a greater rate in the incorporated communities and the proportion of visits by children with severe undernourishment diminished gradually with the length of incorporation of the community.

Progresa/Oportunidades also reduces the likelihood of hospitalization by 2.5 percent (Gutiérrez et al 2005, Gertler and Boyce 2001). The 2001 study shows that hospital inpatient stays of 0-2 year olds fell by more than half and visits to private doctors fell a third. Results were similar for the 18-50 age group, but mixed for 50+ year olds with Gertler and Boyce (2001) reporting large reductions and Gutierrez (2005) reporting an increase of 22 percent for the same age group.³³ When hospitalization occurs, Oportunidades beneficiaries are likely to stay for shorter periods; average length of stay is 1.35 days less than among non-beneficiaries. Although these analyses have not yet clarified whether decreases in hospitalization are a result of more or better utilization of preventive

³³ Nevertheless, the study notes that the statistical capacity to identify differences is limited since hospitalization is an infrequent episode. (Gutiérrez et al 2005).

services or the effects of better knowledge, taken together with the overall increase in preventive care utilization, it appears that the program is successful in reducing the frequency and severity of morbidity (see morbidity section for more detail).

Handa and Huerta (2000) examined an additional hypothesis that Progresa generated community or spillover effects on non-beneficiaries based on anecdotal reports from health workers in Progresa communities that indicated that overall service use had increased due to a demonstration effect (would not be relevant where individual household targeting is not applied). These authors did find that clinic attendance rates for non-beneficiary children living in Progresa communities were five percent higher than non-eligible children living in non-Progresa communities, thus indicating that the beneficiary-only impact assessments are an under-estimate of the full effects of the program.

Oportunidades had mixed effects with respect to the use of pre-natal care. The INSP (2003) study found no significant differences with respect to the proportion of pre-natal care initiated during the first trimester of pregnancy between the three groups of analyzed clinics were found nor an important increase in general in this proportion. Prado et al (2004) find that the percentage of births with appropriate pre-natal care defined as at least five visits during pregnancy increased by 6.12 percent between 2002 and 2003 among urban beneficiaries. Using the rural sample, however, the authors find no impact. The intervention groups were found to seek pre-natal care earlier during the pregnancy, to have a larger number of pre-natal care visits, and a higher proportion of this group have a satisfactory pre-natal care index compared to the control group. No significant effects were found on qualified delivery care (Prado et al 2004).

Beneficiary families in Oportunidades may have changed their health care seeking behavior by substituting public for private services; the mid-term rural evaluation reports that the use of public health services increased in the same proportion that private ones decreased (Gutiérrez et al 2005), possibly indicating substitution from private to public health providers.

For PRAF, IFPRI (2003) reported an increase in the percent of children under age 3 that visited a health clinic in the past 30 days by about 20%, that have had a growth check-up in the last 30 days by 20–22% and that have been weighed in the last 30 days by about 15%. PRAF evaluators also recorded an increase of about 20 percentage points in the percent of pregnant women who had at least five prenatal check-ups. Morris et al (2004) evaluate the impact of PRAF on the use of pre-natal care, vaccination and growth monitoring. Significant effects are found only in the groups that received demand treatments (transfers) and not among households that only benefited from supply-side interventions. Morris et al also report that utilization increases were concentrated among children ages 1-5 years old.

RPS produced a significant average increase of 11 percentage points in the percent of children under age 3 whose parents had taken them for a well child visit in the six months prior to the evaluation in 2002. An average effect of 17.5 additional percentage points was recorded with respect to those children who were taken to a health provider and

weighed (Maluccio and Flores 2004). Average program effects are larger for poorer households. Effects were more pronounced for children ages 3-5 years old. Unlike the Mexican case, where beneficiary families substituted public for private care, the Nicaragua program provided services to beneficiary families through contracted NGO providers. In this setting, some beneficiary families may have substituted private/NGO care for public care, possibly resulting in an increase in utilization among control households as well given that control areas are geographically adjacent and beneficiary-related demand for services shifted.

FA beneficiary children also displayed pronounced increases in growth monitoring visits to health providers; an increase of about 30 percentage points among children under 24 months and 50 percentage points between 24 and 48 months, all highly significant differences (IFS-Econometria-SEI 2004). No differences among those older than 48 months were observed.

While the CCT evaluations allow for an assessment of changes in beneficiary care-seeking practices with respect to total number of visits, particularly well child visits, the analyses published to date provide limited information on a number of the assumptions and causal pathways described earlier that, in turn, limit the interpretation of the findings.

First, the underlying motivation for increased utilization may indeed be related to program conditionality, but it could also be the result of a combination of any number of the factors that the programs influence such as increased parental health knowledge, provision of nutritional supplements in public health centers, increased income, improved nutritional status and health practices which reduce need for service utilization, and/or improvements in supply quality. With the exception of the PRAF design that included four study groups (cash transfers, supply boosts and the combination of cash transfers and supply boosts along with a control group), there has been little effort to distinguish the role of each factor in the utilization changes observed. Unfortunately delays in the implementation and distribution of the supply-side interventions during the evaluation data collection period limited the usefulness of this prescient evaluation design in studying the role of various factors influencing utilization.

Second, net increases in utilization remain difficult to measure given that some programs, such as Colombia and Honduras, do not report on what happens to private sector utilization. In addition, the only experimental evaluation design – Oportunidades – yields no effects on utilization for the youngest age groups.

Third, the age patterns of utilization effects remain difficult to interpret. An intensification of effects is noted in the 2-5 year old group, in spite of the fact that 0-24 month olds are most vulnerable to morbidity and mortality and are thus considered the most critical target population for preventive health interventions. Some hypotheses on this phenomenon have been offered; higher baseline use exists among the youngest age group or better maternal nutritional status combined with increased duration of breastfeeding (not verified where this was reported) may lead to healthier babies and thus lower use of curative services. In general, better exploration of the relationship between

age patterns of utilization effects and age patterns in terms of decreases in morbidity and mortality would help disentangle which parts of the CCT-health package influence impact and whether there is link between service utilization and health outcomes.

Finally, while the increases observed in preventive care visits are inherently positive, in particular when observed together with a decrease in curative care, it is also possible that above a certain amount of visits, crowding out of non-beneficiary clients may take place or the quality may drop to such levels that such increases are no longer desirable under the existing supply constraints. There could also be household welfare costs associated with excess utilization. For example, an average household at the start of Progresa faced 32 conditioned visits for health care and talks a year. Understanding what the social and individual optimal levels of utilization are will be helpful for future program design.

Outputs: Health knowledge, attitudes and practice. Direct measurement of health knowledge and attitudes has generally not been a component of the evaluations (or if included in questionnaires, has not been reported in evaluation reports or papers as in the case of Colombia), in spite of the inclusion of health education components in all programs. As knowledge, attitudes and practices are posited to be critical elements in health outcome changes, understanding the evolution of these factors is critical to explaining the health results observed.

There are a few exceptions to the lack of study of the impacts on health knowledge and behaviors. Progresa evaluators found an increase in dietary quality and calorie consumption. After controlling for the income effect associated with increased calorie consumption, the increase in consumption of more diverse, high nutritional quality foods, such as fruits, vegetables and animal products, indicates that there may be an effect of the nutritional education provided through health education talks known as “prácticas” (Hoddinott et al 2000). This effect did not vary systematically by socio-economic status of the beneficiary families. Duarte et al. (2004) attempts to isolate the effects of the Oportunidades health talks on smoking and alcohol consumption in rural areas, an analysis that is also interesting given that a potential income effect of the transfer is to increase consumption of these goods, finding that youth consumed less alcohol and more cigarettes than control groups. Moreover, Prado et al (2004) report an increased knowledge of family planning methods in both urban and rural areas. In addition, in rural areas use of modern family planning methods is higher in the intervention than in the control group. Breastfeeding initiation and duration was included in many questionnaires, but few reported on results. Surprisingly then, the lecture-style health talks accompanied by peer to peer learning through the “community mothers” seem to have had positive effects in Mexico in some areas.

Evaluators found that the Colombia FA program increased the time that children are breastfed by 1.44 months in urban areas and 0.84 months in rural areas. The program also improved the quality of food consumed by children, increasing the average number of days per week in which various proteins (between 0.25 days and 1.09 days), grains (about 0.45 days) and fruits and vegetables (between 0.45 days and 1.23 days) are consumed. In addition, the program increased overall household consumption of high quality foods -- proteins, tubers, cereals, fruits and vegetables and fats and oils in urban

areas and proteins, cereals and fats and oils in rural areas. More evidence of CCT program impact on improving healthy practices comes from Hoddinott and Yohannes (2002), who find that RPS had a beneficial effect on dietary diversity in terms of both the number and quality of foods purchased.

Outputs: Consumption. Beneficiary communities and households of CCT programs are mostly poor or extremely poor prior to program start; the latter by definition have a total level of household expenditures that is insufficient to purchase a basic basket of foodstuffs. In addition to improved quality of foods consumed, increased food consumption is considered to be the vehicle for the achievement of improved nutritional status, which is linked in turn to better cognitive and social development, higher levels of educational attainment and other outcomes.

Hoddinott et al (2000) find that mean household consumption increased significantly among Progresa beneficiary households, a difference of almost 11 percent, and that this effect was more pronounced among the poor (see Table 7). As reported above, the value of food consumption is also significantly higher as the increase in consumption is concentrated among fruits, vegetables and animal products. There was no evidence that the provision of the fortified food supplement targeted to young children crowded out caloric consumption.

The average effect of RPS on per capita annual household expenditures was, in 2002, about 13 percent of its pre-program value and the effect was much larger for extreme poor households than for the non poor (fourty percent of initial per capita expenditures for extreme poor households versus 6 percent for non-poor households) (Maluccio and Flores 2003). RPS evaluators also reported increases in per capita annual food expenditures, more or less equivalent to the increases observed in total per capita annual household expenditure, reflecting that most of the additional income supplied by the transfer was put towards the purchase of foodstuffs.

For PRAF, IFPRI (2003) reports the results of the first evaluation that was conducted after two years of program operation. Many operational problems delayed the implementation of many components of the program, especially those on the supply-side. As mentioned earlier, although the evaluation was designed to be able to observe impacts separately from demand-side and supply-side interventions, the problems with implementation meant that only the demand-side of the program (cash transfers to households) could be evaluated. However, payments were very irregular, which is believed to have resulted in the very small or non-existent impacts observed. In the case of food consumption, no significant effects were observed.

The Colombia FA program is found to have increased total consumption, consumption of food and expenditures on children's clothing in both rural and urban areas, as well as expenditures/consumption on schooling in urban areas. Specifically, the program increased total consumption by 0.130 log points in urban areas, and by 0.156 log points in rural areas. Moreover, food consumption increased by 0.186 log points in urban areas, while that in rural areas increased 0.203 log points. As reported above, the increased consumption was mainly concentrated among high-quality foods.

Thus, with the exception of the Honduran program, the effect of the program on total household consumption is large and significant, but unsurprising given that CCT programs transfer cash directly to participating families. More importantly, the effects on food consumption and diet diversity are large and significant and the results of this increased caloric quality are evident in some of the nutritional outcomes described below.

Outputs: Supply of health services. The CCT program effect model includes two main assumptions related to supply of health services; first, that current supply of health services is adequate or an increase will follow the increase in demand resulting from the program and, second, that utilization will improve health status. The latter further assumes that the quality of care available is sufficient to effect positive changes in health. While most of the programs have administrative data that could be tapped, there is currently little documentation on how the programs have affected the availability and quality of health care services and, of particular relevance, how supply-side program components have affected the observed outcomes.

As reported above, Gertler and Boyce (2001) record substantially increased numbers of visits in Progreso localities using administrative data, while qualitative studies also confirm increased workloads. In an urban setting, medical staff in beneficiary communities report 23 to 87 percent more visits (Latapí et al 2005), while a focus group of health directors report staff shortages, saturation of services and lack of supplies (Meneses et al 2005). In response to these challenges, some medical staff report charging program beneficiaries when a medical visit is not related to the program. In some cases, medical staff diagnose illness during the program mandatory consultation, but beneficiaries are requested to make an additional appointment and are charged for this “extra consultation” (Latapí et al 2005).

In terms of increases in physical facilities, SEDESOL (2006) reports an increase in the number of health clinics in program localities and in public budgets for health. Latapí et al (2005) report that urban clinics were built in program areas after the program’s introduction and include higher-salaried staff, both of which may indicate a supply response to the program.

As is noted elsewhere, the Nicaragua program directly financed scaled up health supply through NGO providers. Regalia and Castro (2006) discuss the increase in the number of health care facilities due to RPS, but do not discuss issues of quality or the effect of the health lectures on health-related behaviors and knowledge in beneficiary households. While the fact that dedicated funding for increases in health care facilities, staff, training, equipment and supplies should translate directly into improvements in these areas, in practice, delays in implementation (of training, hiring and resource transfers) could result in large deviations between planned and actual changes, especially over the short periods of time common for impact evaluations. Thus, it is not a trivial exercise to evaluate how well and how quickly the planned supply-side resource transfers are executed. Up to this point researchers have been unable to fully separate the effects of the various components of the program, especially the differences between impacts due to the cash transfers versus supply-side improvements. However, without determining the changes that occur in the supply of services, it is impossible to conduct such analysis.

Outputs: Quality of health services. Quality of services gives content to the conditioned health center visits and, without adequate quality, expected effects will not occur. Most of the data on quality comes from small-scale and qualitative studies, so findings cannot be generalized. The few existing studies are Mexican. In spite of these caveats, it is worth noting that while the number of procedures is higher amongst beneficiaries, the results of the interventions are not encouraging, suggesting that supply-side strengthening for quality of care must be a priority.

Availability and quality of medicines appears to be a major issue in Mexico. A small-scale facility survey by Escobar and Gonzalez (2002) found that public health clinics in a group of Oportunidades localities had insufficient supply of medicines available to treat the increased number of patients. Beneficiaries reported that the medicines provided by the public clinics were perceived to be of low quality and that many beneficiaries were choosing to purchase more expensive higher quality drugs at private pharmacies. Neufeld et al (2005) noted that delivery of nutritional supplements to program localities in Mexico was sometimes delayed, resulting in inadequate supply and potentially reducing the frequency of consumption.

A study on quality of care for Oportunidades diabetic beneficiaries in urban areas (Hernandez-Avila et al 2004) attempts to link quality of care to observed differences between treated and untreated households and individuals. Researchers collected blood samples from diabetic patients and recorded the extent of compliance with diabetes treatment protocols by both providers and patients. Findings indicate that diagnoses of diabetes have increased, but beneficiaries are no more likely to have the disease properly treated, suggesting that there is much room to improve the quality of health care that beneficiaries receive. Using the full sample, Fernald et al (2004) confirm that beneficiaries are more likely to be checked for diabetes, although no significant differences are observed in terms of the effects of treatment.

Another Mexican study on the rural Oportunidades sample (Prado et al 2004) examines the quality of pre-natal and delivery care and finds that beneficiary women in the intervention group received on average a larger number of the procedures stipulated in the protocol. However, the frequency with which the women were informed of their blood groups or were administered syphilis detection test was very low. No significant differences were observed between the groups with respect to births in a medical facility, although a smaller proportion of cesareans were recorded in the early intervention group than in the control groups.

The FA impact evaluation included a health facility survey which collected information about various characteristics related to access and quality of care. The health facility surveys collected information about the hours of operation, types of services offered, the number of various services provided in the past year, number and type of current staff, stocks of various medicines, interruptions in service due to labor problems, political unrest or natural disasters, the previous year's budget and revenue, main sources of revenue, participation in and training received for the program and other details related to the program. However, the sampling method of facilities is not clear and this information was not studied or included in the impact evaluation.

Output/Outcome: Vaccination rates. Vaccination protocols differ slightly between countries, but a “full scheme for age” according to the Expanded Program on Immunization generally includes by exact age 1 – one dose of BCG, three doses of polio, three doses of pentavalent or DPT and one dose of MMR. Although vaccination was a condition for receipt of transfers in Honduras and Nicaragua, in practice, this condition was not monitored by the CCT program. Instead only the visit was recorded, rather than the immunization schedule completion progress, though it is frequently the practice of health centers to maintain immunization records as well. As a result, estimates of the impact on vaccination rates were done through the external evaluation of the programs. The evaluation of rates is complicated due to poor recording practices in family-held vaccination cards on the part of health providers. All of the evaluations report difficulties in establishing the date of vaccination. Some evaluations look at only one or two “tracer” vaccines, while others attempt to document full vaccination (though the definition used by each evaluation is inconsistent).

The overall contribution of CCT *per se* to vaccination coverage appears marginal. In spite of apparent program-attributable increases during a pilot implemented during 2000-2001, the Nicaragua RPS produced an insignificant average net increase of 6.1 percentage points in up-to-date vaccination levels between 2000 and 2002 (Maluccio and Flores 2004).³⁴ PRAF children showed marginally higher rates of DTP/Pentavalent, though insignificant and small differences for MMR (Morris et al 2005). IFPRI (2003) reported an increase of 8 percent in children under age three that received their first dose of DPT on time. The Colombia program measured DPT prevalence and found an insignificant difference between program participants versus controls (IFS-Econometria-SEI 2005).

Vaccination is difficult to impose as a condition as it is a very supply-dependent service; unlike growth monitoring visits, if vaccines are not in stock, vaccination will not occur. The Honduras experience, where supply was variable, seems to be an example of this phenomenon and results may relate more to the availability of vaccines at health centers than a demand effect, although there is no way to assess this empirically. However, there may be an indirect effect of the program in that coordination with the MOH in program areas may in fact generate more vaccine supply (Maluccio and Flores 2004).

Outcomes: Nutritional status. Unlike the unambiguously positive results with respect to food consumption in most programs, nutritional status outcome measures show a mixed result (Table 6). Leaving aside the Honduras program, all of the programs show a significant reduction in stunting. Results for the proportion underweight are less consistent; Nicaragua RPS shows a large and significant decline, while the Colombia program shows impact only in rural areas for 3-7 year olds. In Mexico, no impact is observed on this measure. No effects on anemia are observed, despite the evidence of greater diet diversity from the consumption data.

³⁴ Maluccio and Flores (2004) include an interesting footnote regarding the quality of administrative data on vaccination that will be relevant to other payment for performance schemes; they find that survey reports are substantially lower than the 100 percent recorded in administrative data. The errors may go in both directions. (see footnote 30 – and there could be an entire separate paper on this).

After a year of implementation, Progresa was reported to increase the growth of children between 12 and 36 months old by about one centimeter per year, which is approximately one-sixth of the growth which would have occurred without the program (Behrman and Hoddinott 2000). Weight at birth increased in all the groups, but the increase was higher in the eligible women from no intervention zones than in the incorporated group (i.e. the proportion of births with low weight at birth decreased more significantly in the former).

Although early evaluations found no impact in anemia, Gertler and Fernald (2004) find that after 6 years of operation, there was a small but insignificant decrease of about 3 percent in anemia overall. However, the magnitude of the decrease varied by the age at which the program began for a child, with a larger decrease (of 26 percent) for those whose treatment began in the antenatal period among children aged 24-72 months old at the time of the 2003 survey. By the time of the 2003 survey round used in this study, all treatment children had been on the program for more than two years, and about half the sample had been receiving benefits since conception. This same study indicates that among children 24-72 months old, a reduction of stunting of twenty nine percent is found for girls (58 percent if treatment started in the antenatal period), and eleven percent for boys (16 percent for the antenatal). One concern is that the program appears to be significantly increasing obesity among boys by about 40 percent (and among girls by fifteen percent). Both treatment boys and girls are significantly less stunted, however treatment boys tend to be more overweight and obese.

RPS shows the strongest and most consistent results for reductions in stunting and weight-for-age; on average, evaluators observed a 5.5 percentage point decline in the percentage of children who were stunted (height for age), from an initial level of 39% (Maluccio and Flores 2005). RPS also conducted a qualitative evaluation that allows for better development of explanatory hypotheses for many of the results recorded in the quantitative evaluation. Although the program included iron supplementation in its visit protocols and launched an aggressive program of distribution, the program had no impact on anemia rates. The qualitative evaluation found that mothers knew that supplements were important for their children's health, yet were not regularly administering these supplements to their children for a variety of reasons, including the fact that children did not like the taste or supplements caused vomit or diarrhea (Adato and Roopnaraine 2004).

As noted above, operational problems in PRAF are hypothesized to result in no recorded effects on height-for-age, weight-for-height and weight-for-age or the incidence of iron-deficient anemia.

As is expected from the increased consumption of high nutrition foods among children, the FA program is also found to have increased height-for-age and weight-for-age z-scores among children under 24 months old by 0.161 and 0.192, respectively. In addition, the program improved weight-for-age among children over 48 months old, increasing the z-score by 0.107. These improvements in the height-for-age and weight-for-age measures translate into a reduction in the proportion of children less than 24 months old classified as stunted by 6.9 percentage points and a reduction in the proportion of children over 48 months old classified as stunted or at risk of undernourishment by 3.7 percentage points. Separating the effect by urban and rural

regions, the program is found to have reduced the proportion of children over 36 months old in rural areas classified as underweight by 3.4 percentage points and the proportion of children less than 36 months old in urban areas that are underweight or at risk of undernourishment by 14.2 percentage points. FA is found to have increased birth weight in urban areas by 0.578 kilograms, but to have had no effect on birth weights in rural areas or on the prevalence of underweight or overweight among women with children under age 7. In addition, the program is found to have reduced the incidence of diarrhea in the past 15 days among children under age 4 in rural areas by about 10 percentage points.

An exception to the positive effects of CCT programs on child nutritional status is reported by Morris et al (2004) for Bolsa Alimentacao (a pre-cursor to Bolsa Familia) in Brazil. The authors examine data from northeast Brazil to assess the impact of the program on the rate of child weight gain (through weight-for-age measurements of children under age seven), finding that the average weight-for-age z-score was 0.13 lower among beneficiaries than among excluded children. Using monthly growth measurements for children under age 3, it was determined that each month of exposure to the program reduced weight gain by 31 grams. The authors believe that the negative effect of the nutrition program is due to the fact that mothers thought that if their child began to grow, they would be removed from the program, and thus had the incentive to reduce the growth of their child or to avoid providing the added nutrition provided by the program. The results are only valid for the pilot phase and cannot be assumed to hold for the larger expanded program.

Outcomes: Morbidity. Measured morbidity may increase or decrease as a result of the CCT intervention. On the one hand, greater use of preventive care and higher levels of health knowledge may lead to fewer episodes of illness. On the other hand, greater levels of health knowledge and more frequent visits to health centers may increase the probability that mothers more frequently and/or accurately diagnose illnesses and seek care when it is required. Moreover, improved health may lead to increased symptoms of morbidity (such as respiratory infections) which cause an increase in the demand for curative care (Oppenheimer 2001). These hypotheses have not been much explored by the evaluations and most results that are available come from the Mexican program. Where it has been measured, CCT appear to decrease the incidence and prevalence of morbidity.

The Mexico evaluation found that the impact of the program on the probability of child illness is negative and statistically significant for all age groups, but not until a child had been receiving benefits for at least a 12 month period (Gertler and Boyce 2001). In rural areas, a small decrease in sick days for only the productive age population is observed, while a larger effect and wider age range is seen in urban areas. A later study by Gutierrez et al 2005 finds that the number of days ill during the four weeks prior to the survey fell by about 21 to 24 percent for boys and 9 to 10 percent for girls. Overall, the number of days lost to illness decreased by 20% among beneficiary families and have experienced a decline of slightly more than 1 day in the number of days of short-term disability.

Gutierrez et al (2004) use data from the Progesa/Oportunidades rural and urban evaluation samples, ENCEL (2000, 2003) and ENCELURB (2002, 2003) to determine the impact of Oportunidades on sick days and physical capacity. In rural areas, the only significant difference between the intervention and control groups for the number of sick days was found for the 16-49 years group (i.e. the productive age) – a decrease in sick days of 2.1 per year. There were no significant results for disability days, although the effects were negative (decreasing). As for the capacity to carry out physical activities, the 18-49 years group experienced a significant, although small, increase (0.5 %). No significant effect on number of kilometers or blocks that the person can walk. In urban areas, the only significant difference between the intervention and control groups for the number of sick days was found for the 6 - 15 years group – a decrease in sick days of 8.3 per year. As for 16- 49 years group, although not significant, the effect is the equivalent of a decrease in sick days of 6.8 per year. As for disability days, the results were significant and equivalent of a reduction of disability days of 6.2 per year for the 6-15 years group, and 7.57 per year for the 16-49 years group. As for the capacity to carry out physical activities, small positive effects were found. An anomalous finding is that the paper shows a larger decreasing effect on reported illness and disability in the urban group than in the rural (larger effect and for a wider age-range), in spite of the fact that there is a larger program effect on utilization of services in the rural area.

An Oportunidades study of the effect of the program on indigenous populations found that the program decreased illness rates from 0.2 to 3.5 percent in the three age ranges studied (0-5, 3-5, and under 3), with the greatest effects observed among children under three years old (Quiñones 2006). The average program effect estimations differ substantially from Gertler's findings, most likely due to differences in sampling size and approach used. Positive program effects observed for indigenous beneficiaries were similar to those of non-indigenous beneficiaries. This finding could be considered a weakness of the program, as is argued by Quiñones, given the implicit assumption that the average effect of the program should be larger on poorer and more marginalized groups. On the other hand, it is remarkable that the benefits experienced are equivalent for both groups when one takes into account anecdotal evidence that indigenous groups have a more difficult time complying with CCT conditionalities due to language and cultural barriers.

In terms of diarrheal disease, IFPRI (2003) reports that PRAF generated a reduction in the incidence of diarrhea among children under age three between three and six percent. Colombia FA evaluators find that the program reduced the incidence of diarrhea in the past 15 days among children under age 4 in rural areas by about ten percentage points. These results may be explained by better nutritional status, better hygiene practices or better vaccination coverage.

With respect to chronic disease-related morbidity, the required regular check ups and participation in health talks may have a positive effect on household and social norms related food intake and activity. On the other hand, the income transfer to the household could make some behaviors that increase the risk of chronic diseases more affordable (e.g. junk foods, soft drinks, alcohol). The single study on this issue from Mexico finds encouraging results. Fernald et al (2004) find high baseline prevalence of obesity (twenty

four percent), hypertension (thirty nine percent) and diabetes (nineteen percent) among the rural poor in Mexico and that Oportunidades participation significantly reduces the prevalence of all except diabetes. Symptoms of hypertension and diabetes are also significantly reduced via program participation. Note that the Mexico program is the only to include the entire household as eligible beneficiaries for the CCT program; these evaluation results suggest that if a CCT is implemented to achieve health objectives in countries well into the epidemiological transition, requiring poor adults to seek preventive care and check ups may be an effective strategy. Combined with the qualitative information on quality of care, however, another hypothesis worth considering is that adult health outcome improvements can be achieved by providing beneficiaries with information about their chronic disease status in combination with nutrition education and a cash transfer.

Outcomes: Mortality. Mortality was not measured directly by the evaluations; only the Mexican program uses administrative data to analyze program effects. While the Mexico program finds that infant and maternal mortality declined significantly in program areas, the reliance on administrative data, particularly in the case of maternal mortality, is problematic.

Hernández et al (2004) examine the impact of Oportunidades on maternal and infant mortality using data for the period 1995 – 2002 from the Ministry of Health (SSA) and the National Institute of Statistics, Geography and IT (INEGI), disaggregated at municipal level. For the entire period, once adjusting for the effect of the year, the presence of other programs, number of health units and marginalization, maternal mortality was 11 percent lower in the municipalities with at least one locality incorporated in Oportunidades compared to those without any localities incorporated. Estimates of absolute numbers indicate that an average of 79 maternal deaths per year have been avoided in this period thanks to the program. The reduction in maternal mortality is larger in municipalities that have between 4 percent and 35 percent of the population incorporated in the program, reaching a 15 percent reduction.

An interesting aspect of the intervention package and the measured impacts is that the single most effective preventive visit in terms of child and maternal mortality, institutionally attended births (or Emergency Obstetric Care³⁵), is not a condition in any of the programs (except for PRAF, where the condition has yet not been implemented). Most life-threatening complications occur during labor and delivery, and pre-natal screening does not identify all of the women who will develop complications. In fact, most obstetric complications occur among women with no risk factors (Rooks, Winikoff, and Bruce 1990). As noted earlier, there was no measured effect of the OPORTUNIDADES program on institutional births. This finding, coupled with the relatively imprecise methodology applied for measuring program effects (why would you find a significant effect in a large municipality with only one incorporated community, and why would the effects not increase linearly with the amount of communities in a

³⁵ EOC includes specific interventions to manage "emergency" obstetric complications. Interventions may be intravenous antibiotics, oxytocics or anti-convulsants, management of abortion complications, management of postpartum bleeding, assisted delivery for prolonged labor such as vacuum or forceps delivery, blood transfusion, and/or cesarean section.

given municipality that are actually incorporated in the program?) calls into question the reliability of the result.

For the entire period (1997 – 2003?), infant mortality was two percent lower in the municipalities incorporated in Oportunidades compared to the non-incorporated ones. Infant mortality showed a generally decreasing tendency during the period, independently of incorporation into Oportunidades. Estimates of absolute numbers indicate that an average of 340 infant deaths per year have been avoided in this period thanks to the program. The impact of Oportunidades on infant mortality at the municipal level increases as the proportion of the population incorporated in the program increases - reaching a reduction of six percent in municipalities with more than 35 percent of the population participating.

Outcomes: Fertility. While the literature on financial incentives in fertility suggests that fertility is in a secular decline throughout the developing world and that welfare programs and state policies have not been sufficient to generate a positive fertility response, there was a concern among some that fertility levels might be affected by CCT programs. Although Colombia, Mexico and Nicaragua data indicate that fertility rates decreased in the presence of the program, the Honduras program, which applied a different payment incentive structure, observed a fertility increase.

Stecklov et al (2006) have examined the unintended effects of CCT programs in Honduras, Mexico and Nicaragua on fertility levels, finding that unintentional incentives for childbearing in Honduras (a health/nutrition subsidy that is not lump sum and varied by number of children and pregnant women in a beneficiary household) may have contributed to a 2-4 percentage point increase in fertility. This effect, which was not observed in Mexico and Nicaragua where health/nutrition grants are lump sum, may be related to an increase in marriage rates, the effects of the program on the presence of the partner, or a tempo response to the program's unintended incentives.

Prado et al (2004) find that in rural areas, no difference was observed between Oportunidades intervention and control localities in terms of use of family planning methods, although the proportion of women using these methods decreased in both groups. Nevertheless, the average number of children per woman in reproductive age also decreased in both groups.

The Colombia FA evaluation reports a relatively large decline in fertility across the board (control and intervention groups) between baseline and follow up surveys, but the reasons behind these changes are not explored in the evaluation report.

Evaluation limitations

While the CCT program evaluations certainly set a high standard for impact evaluation, there are nonetheless major limitations related to the sampling designs and construction of counterfactual groups. As these issues are primarily methodological, they have been included in an annex. With respect to health in particular, the limitations of the CCT evaluations relate to the use of instruments and particular questions appropriate for

studying the relationship between specific components of the program and specific outputs or outcomes. Aspects in which study has been limited and for which improvements could enhance both the knowledge of program impacts and the design of future programs are highlighted and discussed below.

CCT program evaluations have given very limited attention to the impact on health-related behaviors, attitudes and household decision-making or how these factors contribute or limit impacts on outcomes. As is clear from the previous summary, the majority of the program evaluations have focused on measuring the effects on specific health indicator outcomes, such as incidences of illnesses, child growth and outputs such as utilization rates of public facilities for preventative, curative and prenatal care. Granted, improvements in these outcomes and outputs were listed as goals by most programs, so determining the program's impact is an important first step. However, many of the outputs, especially use of preventive and prenatal care are actually conditions for receipt of transfers, so it should be expected that the program should lead to increases. Moreover, outcomes such as illnesses and child growth are observed more or less at the end of a black box, with evaluations providing little understanding of exactly how the program operates to effect such changes in outcomes.

For example, both Oportunidades and RPS were found to have improved health outcomes among young children, but it is not clear exactly which components were important in achieving the improvements. For example, these impacts could be due to the receipt of nutritional supplements, increased use of preventative and prenatal care services, increased food consumption, increased knowledge in topics covered by the health information lectures (such as proper hygiene and food preparation, best practices for breastfeeding and treatment of diarrhea) or even increased coverage and timeliness of vaccinations through increased supply of preventative care services. Understanding the role of such factors in influencing outcomes is critical for developing more effective programs. For example, a rare and recent exception to this gap in the evaluations is the evidence that has emerged about the limited role of the nutritional supplements in improving child growth and nutrition in Oportunidades. Neufeld et al (2005) studied the proper preparation and consumption of the nutrition supplements by children under five in urban beneficiary households. Their findings suggest that the supplements could have a greater impact if a greater quantity was provided to families (because doses are commonly reduced because sharing with non-targeted household members is common) and that more attention should be paid to ensuring on-time delivery of the supplements to localities and in how the information about proper preparation is delivered to beneficiaries. In addition, the form of some nutrients was found to be less easily absorbed, leading the program to modify the formulation (IDB 2005). Thus, without careful study of the delivery process and use of the supplements by beneficiaries, these issues would continue to limit the effectiveness of at least one component of the program.

The fact that the nutrition supplements were often not prepared and consumed properly by beneficiaries in Mexico suggests that the growth improvements observed among young children were mainly due to increases in the quantity and variety of household food consumption. However, a key component missing in this analysis is an understanding of how well other information presented in the lectures was received and

incorporated into beneficiary households. Qualitative studies have indicated that the information transfer in the health lectures was minimal, in part due to the fact that they were at first only delivered in Spanish and not in native languages. However, behavioral changes could also be studied quantitatively through household surveys that collect detailed information about breast and supplemental feeding practices of young children, treatment of incidences of diarrhea and other illnesses (not just documenting occurrences), immunization records for young children, outcomes of current and past pregnancies (not just prenatal care) and other information related to the content of health lectures. The Demographic and Health Surveys (DHS) can provide a good model from which to base the design of future evaluations surveys as their questions have been tested repeatedly in a variety of social and cultural settings.

Finally, although the program rationale and effect model indicate that the reduction of out of pocket and opportunity costs associated with health care seeking is both the principal mechanism to increase utilization and thus one of the outcome variables on which one would expect to observe impact, no evaluation has analyzed these aspects. Many other examples can be cited.