# EVALUATION OF ANTI-POVERTY PROGRAMS' IMPACT ON JOINT DISADVANTAGES: INSIGHTS FROM THE PHILIPPINE EXPERIENCE

BY SUMAN SETH*

*Leeds University Business School*

AND

MELBA VERRA TUTOR

*Independent researcher*

Anti-poverty programs increasingly target multiple outcomes to address current and future poverty. Conventional evaluation exercises, however, estimate programs' impact on outcomes separately. We present a framework, drawing from the counting approach, that captures the joint distribution of outcomes and allows evaluating program impact on the distribution of multiple outcomes. We apply the framework to analyze the Philippine conditional cash transfer program using an embedded randomized control trial survey. We observe that the program induced targeted behavioral changes by reducing overall noncompliances to conditionalities, but did not necessarily benefit the poorer families experiencing a larger number of simultaneous noncompliances. The overall positive impact is achieved by leaving the neediest behind. Our results show the valuable contribution of considering the joint distribution of outcomes in evaluating poverty alleviation programs.

## 1. INTRODUCTION

Poverty alleviation strategies and programs are a fundamental component of welfare policies in both developed and developing countries. They range from a variety of welfare programs in the US (e.g. see Council of Economic

Advisers, 2018) and strategies for tackling poverty, social exclusion, and social immobility in the European countries (OECD, 2007, 2018) to social security programs enhancing food and livelihood security in India (Dutta et al., 2014) and a multitude of social safety net programs in developing countries across the globe, which include cash transfers, in-kind transfers, social pensions, food security, livelihood security, and feeding programs targeted to poorer sections of the population.

Many anti-poverty programs by design target disadvantages on multiple outcomes simultaneously. Typically, program *theories of change* rest on addressing simultaneous disadvantages to break the inter-generational cycle of poverty. For instance, conditional cash transfer (CCT) programs aim to tide over families from chronic hunger (present poverty), while simultaneously incentivizing access to schooling and health care to arrest future poverty (Fiszbein and Schady, 2009). Yet, conventional program evaluation exercises often examine the impact of a program on different outcomes separately and determine that a program is successful if it improves some, if not all targeted outcomes. Although these exercises are informative, they do not tell us whether program-induced changes benefited those who were initially facing fewer disadvantages or those who were initially facing a larger number of disadvantages.

In this paper, we aim to make two key contributions to the literature on program evaluation. First, we recast the *counting approach framework* as an evaluation tool to study program impacts on joint multiple outcomes. Second, we evaluate the impact of a large CCT program on targeted outcomes and on multidimensional poverty, showing that the framework generates findings and insights that are missed by single-outcome evaluation exercises.

The counting framework is particularly useful when the underlying indicators take binary forms, i.e. when each indicator can be categorized into those that satisfy an outcome criterion versus those that do not. It is used for developing various social metrics, such as measures of social exclusion (Chakravarty and D'Ambrosio, 2006), chronic poverty (Foster, 2009), multidimensional poverty (Atkinson, 2003), Alkire and Foster (2011), Bossert et al. (2013), women empowerment (Alkire et al., 2013), and vulnerability (Dutta and Mishra, 2018). Among these metrics, the multidimensional poverty measures proposed by Alkire and Foster (2011) are widely applied (United Nations Development Programme, 2010; Alkire and Santos, 2010; World Bank, 2018), and there is also a constructive debate surrounding their applications (Ravallion, 2011; Ferreira, 2011; Alkire et al., 2011).

Although a number of studies have used multidimensional poverty measures to evaluate programs,[1] these studies do not formally present how a counting frame-

---

[1] For example, Robano and Smith (2013) study the impact of the BRAC ultra poverty program on beneficiaries' multidimensional poverty in Bangladesh; Azevedo and Robles (2013) compare a multidimensional targeting approach to the traditional income-based approach for the Mexico's opportunidades program; Loschmann et al. (2015) examine whether shelter assistance in Afghanistan reduces multidimensional poverty; Pasha (2016) studies the impact of cash grants in South Africa on multidimensional poverty; and Song and Imai (2019) evaluate the short-term impact and long-term sustainability of Kenya's Hunger Safety Net Programme on multidimensional poverty. For applications of counting approaches to poverty targeting and measurement, see Alkire *et al.* (2015, Chapter 4).

work can be applied as an evaluation tool to look at a program's impact on multiple outcomes jointly, where these outcomes may account for behavioral changes, deprivations, or any other welfare improvements targeted by the program. In Section 2, we formally show that the counting framework is an effective impact evaluation instrument in capturing changes to the joint distribution of disadvantages. It uncovers program effects on simultaneous multiple disadvantages while still allows for analyzing changes to individual outcomes typical of conventional evaluation exercises.

We then use the framework to study the impact of the Philippine CCT program. CCTs provide *cash grants* to beneficiary families conditional on compliance with prespecified human capital investments, aimed at inducing targeted behavioral changes (Das et al., 2005). CCTs have gained enormous popularity in recent decades as a key social development intervention (Fiszbein and Schady, 2009; Filmer and Schady, 2011; Baird et al., 2011; Glassman et al., 2013; Evans and Popova, 2017; García and Saavedra, 2017).[2] With 4.6 million beneficiaries, the Philippine CCT program is one of the largest in the world. The Philippine government considers it to be a major contributor to recent poverty reduction, and in 2019, the program was institutionalized as the country's flagship poverty reduction program through the "Pantawid Pamilyang Pilipino Program (4Ps) Act."

Studies show that the 4Ps improved outcomes—such as school enrolment, nutritional status, consumption of food and non-food items, and spending for education—and reduced noncompliance rates in different indicators (Onishi et al., 2013a; Onishi et al., 2013b; Orbeta et al., 2014). However, all studies so far look at the program's impact on outcomes and noncompliances separately. To enrich our current understanding, we examine the 4Ps' impact on the joint distribution of outcomes through two separate exercises using a *randomized control trial* household survey specifically designed to capture the impact of the program.

Our first evaluation exercise focuses on the binary changes between noncompliances and compliances among program-specified conditionalities and examines whether the cash grants induced behavioral changes among beneficiaries by reducing their joint *noncompliances*. This exercise looks at the direct intended impact of the program. We capture the program's impact on joint noncompliances by selecting five indicators that directly correspond with the 4Ps conditionalities and by constructing a *multiple noncompliance score*, a sum of the beneficiaries' noncompliances in the selected indicators.

We observe considerable reductions in the incidences of noncompliances (i.e., positive impact) in three of the five selected indicators (between 7.8 and 11.8 percentage points), confirming targeted behavioral changes in these indicators. However, changes in the distribution of multiple noncompliances reveal unsatisfactory results. Although the overall average multiple noncompliance score improved (a reduction of around 0.051 points on a 0–1 scale), there are no significant improvements among families experiencing four or more noncompliances.

Our second exercise investigates whether the program reduced multiple deprivations among beneficiary households. This exercise, which complements the

---

[2]For critical evaluations of CCTs, see Baird et al. (2011), Filmer and Schady (2011), and de Janvry and Sadoulet (2006).

first one, examines the indirect effect of the program using indicators that are *not* directly conditioned on by the program but still capture different forms of deprivations. Similarly, we construct a *multiple deprivation score* by counting deprivations in the selected indicators. Akin to the impact on joint noncompliances, the 4Ps reduced the overall average multiple deprivation score (around 0.024 points on a 0–1 scale), and it also improved—albeit in smaller magnitude—the distribution of multiple deprivations.

Finally, we explore whether the families with more noncompliances are poorer and whether the overall reduction in multiple deprivations is shared by the poorest. We find that families with four or more noncompliances experience more deprivations, on average, than the rest of the beneficiaries. Their consumption and savings behavior show some progress, but apart from these, we do not observe a strong pro-poorest improvement in the joint distribution of deprivations. Thus, despite the program's success in improving various outcomes and noncompliances, the program could not be considered inclusive when we look at the joint distribution of disadvantages, at least during the early stage covered by the evaluation period.

The Sustainable Development Goals aim to reduce poverty *in all its dimensions* by year 2030. The heightened emphasis on interconnected solutions to poverty must correspondingly give rise to evaluation tools that capture the joint distribution of disadvantages. The multidimensional evaluation framework we formalized shows how we can enrich the evidence that we build on anti-poverty programs' effect on simultaneous disadvantages. Our empirical application highlights enhancements to the design and implementation of anti-poverty programs that were not apparent from existing evaluations.

The rest of our paper is organized as follows. We present the counting framework for program evaluation in Section 2. We then present the overview of the Philippine CCT program in Section 3. In Sections 4 and 5, we present the results from our two evaluation exercises, respectively. We explore the relationship between the distributions of noncompliances and deprivations in Section 6. We provide concluding remarks in Section 7.

## 2. Impact Evaluation from a Multidimensional Perspective

We use the term *disadvantage* to refer to the inability to satisfy a welfare criterion addressed by a program. For instance, a disadvantage may refer to a *noncompliance*, which reflects a failure to satisfy a program-specified condition or a *deprivation*, which reflects the inability to meet a minimum requirement of well-being. Let us illustrate how assessing impact on different indicators separately precludes understanding a program's impact on those that are disadvantaged in multiple indicators simultaneously.

$$X = \begin{bmatrix} \text{ND} & \text{ND} & \text{ND} \\ \text{ND} & \text{D} & \text{ND} \\ \text{D} & \text{ND} & \text{D} \\ \text{D} & \text{D} & \text{D} \end{bmatrix} \qquad \overline{X}_1 = \begin{bmatrix} \text{ND} & \text{ND} & \text{ND} \\ \text{ND} & \text{D} & \text{ND} \\ \text{D} & \text{ND} & \text{D} \\ \text{ND} & \text{ND} & \text{ND} \end{bmatrix} \qquad \overline{X}_2 = \begin{bmatrix} \text{ND} & \text{ND} & \text{ND} \\ \text{ND} & \text{ND} & \text{ND} \\ \text{ND} & \text{ND} & \text{ND} \\ \text{D} & \text{D} & \text{D} \end{bmatrix}$$

Suppose an anti-poverty program directly targets three indicators. Let the three matrices—$X$, $\overline{X}_1$, and $\overline{X}_2$—summarize the disadvantage profiles of four *units*, which may represent individuals or households. In each matrix, a row summarizes the disadvantage profile of a unit in three indicators, whereas a column summarizes the disadvantage profile of all units in an indicator. If a unit fails to meet a minimum requirement, the unit *experiences a disadvantage* ('D') in that indicator and thus requires intervention. Otherwise, the unit *does not experience any disadvantage* ('ND') in that indicator.

Before the program (i.e., in $X$), within each indicator (i.e. within each column) two out of four units experience disadvantages. After the program, one of the two alternative disadvantage profiles, $\overline{X}_1$ and $\overline{X}_2$, may be obtained from $X$. The incidence or the proportion of units experiencing disadvantage within each indicator is now a quarter, both in $\overline{X}_1$ and $\overline{X}_2$. Thus, if the impact is evaluated for each indicator separately, then the program appears to be equally effective, whether $\overline{X}_1$ or $\overline{X}_2$ is obtained from $X$.

The difference between the two post-program profiles manifests only when we evaluate the program's impact by considering the three indicators together. In $X$, the first unit does not experience any disadvantage, the second unit experiences disadvantage in one indicator, the third unit in two indicators, and the fourth unit in all three indicators. Now, $\overline{X}_1$ is obtained from $X$ by eliminating all three disadvantages of the fourth unit, whereas $\overline{X}_2$ is obtained from $X$ by eliminating the disadvantages of the other two units, and leaving the fourth unit unchanged. Thus, there may be improvement in each indicator because of the program on average, but it may leave out those with simultaneous disadvantages in a larger number of indicators—those that should, in fact, be prioritized by the program.

To effectively evaluate a program's impact on multiple disadvantages, we present a framework drawing from the counting approach (Atkinson, 2003; Alkire and Foster, 2011). Suppose a program directly targets $d \geq 2$ indicators and the target population contains *n units*. Each indicator, by program design, has a *disadvantage cut-off*. When a unit (denoted by $i$) fails to meet the disadvantage cut-off of an indicator (denoted by $j$), then *unit $i$ experiences disadvantage in indicator $j$ and is assigned a binary *disadvantage status score* of $g_{ij} = 1$. A score of $g_{ij} = 0$ is assigned otherwise. In $X$, $\overline{X}_1$ and $\overline{X}_2$, for instance, a unit is assigned a score of 1 for a status of "D" and 0 for a status of "ND."

The magnitude of multiple disadvantages of a unit is reflected by simply counting its number of disadvantages. A *multiple disadvantage score* (MDS) for unit $i$, denoted by $c_i$, is obtained as $c_i = \sum_{j=1}^{d} g_{ij}$.[3] Clearly, $c_i$ ranges between 0 and $d$ for all $i$ and a higher MDS reflects a larger magnitude of disadvantages. An MDS of $c_i = 0$ means that unit $i$ does not experience disadvantage in any indicator, whereas an MDS of $c_i = d$ means that unit $i$ simultaneously experiences all $d$ disadvantages.

A program evaluator may be interested in evaluating the program's impact on those that experience $k$ or more disadvantages simultaneously (i.e., $c_i \geq k$). We may refer to $k$ as a *disadvantage threshold*, which may be determined by the evaluator's

[3]One may consider different disadvantages of unequal importance and weight disadvantages accordingly. This approach is common in the multidimensional evaluation of well-being and poverty. In the case of program evaluation, an evaluator can also incorporate weights to account for the welfare implications of different program interventions.

normative judgment or the implementing agency's target. For instance, if they aim to capture the impact among all, i.e. those experiencing even one disadvantage, then the threshold should be set at $k = 1$. In contrast, a higher threshold is appropriate when the objective is to evaluate the program's impact on those experiencing a larger number of multiple disadvantages.

A straightforward evaluation exercise is to estimate the change in the *incidence of multiple disadvantages* or the incidence of experiencing $k$ or more disadvantages. Let us denote the incidence of multiple disadvantages for a given disadvantage threshold $k$ by:

$$(1) \qquad H(k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[c_i \geq k] = \frac{q_k}{n},$$

where $\mathbb{I}[c_i \geq k]$ is an indicator function with a value of 1 for $c_i \geq k$ and 0 otherwise, and $q_k$ is the number of units experiencing $k$ or more disadvantages. Clearly, the incidence is bounded between 0 and 1. A reduction in $H$ reflects a *positive* program impact.

An impact evaluation exercise based only on comparing incidences, however, ignores any change in the *intensity* or multiplicity of disadvantages among those that experience $k$ or more disadvantages.[4] A simple way to reflect the intensity of multiple disadvantages may be to look at the average MDS of those experiencing $k$ or more disadvantages:

$$(2) \qquad A(k) = \frac{1}{q_k} \sum_{i=1}^{q_k} \left[ c_i \,|\, c_i \geq k \right].$$

By construction, $A$ is bounded between $k$ and $d$. The lower bound, $k$, is reached when all $q_k$ units experience exactly $k$ disadvantages. The upper bound, $d$, is reached either (a) when all $q_k$ units experience $d$ disadvantages simultaneously, or (b) when we are interested in those that experience all $d$ disadvantages (i.e. $k = d$) and one or more units have such an experience.

Let us illustrate why an impact evaluation exercise should include intensity in addition to the incidence of multiple disadvantages. Suppose, deprivation profile $\overline{X}_3$ is obtained from $X$ by alleviating one disadvantage for the fourth unit. A policy evaluation exercise that focuses on those with two or more disadvantages (i.e. $k = 2$) would reveal no program impact if the exercise merely compares incidences, because two units in both $X$ and $\overline{X}_3$ experience two or more disadvantages. Yet, the intensity of those experiencing two or more disadvantages (third and fourth units) decreased from 2.5 in $X$ to 2 in $\overline{X}_3$. The program did not reduce the incidence of two or more disadvantages, but it commendably reduced one disadvantage for the unit in greatest need of attention.

$$X = \begin{bmatrix} ND & ND & ND \\ ND & D & ND \\ D & ND & D \\ D & D & D \end{bmatrix} \qquad \overline{X}_3 = \begin{bmatrix} ND & D & ND \\ ND & D & ND \\ D & ND & D \\ D & ND & D \end{bmatrix}$$

[4]It is equivalent to violating the dimensional monotonicity property in Alkire and Foster (2011).

Changes in both incidence and intensity of multiple disadvantages may be captured by the following measure, motivated by the *adjusted head count ratio* (Alkire and Foster, 2011):

$$(3) \qquad M(k) = H(k) \times \frac{A(k)}{d} = \frac{1}{nd} \sum_{i=1}^{n} c_i \times \mathbb{I}[c_i \geq k],$$

where $\mathbb{I}[c_i \geq k]$ is an indicator function. Measure $M$ is a product of both incidence and intensity of multiple disadvantages divided by the number of indicators. The maximum feasible number of disadvantages is $n \times d$, which occurs when all $n$ units experience $d$ disadvantages. The minimum feasible number of disadvantages is zero, which occurs when every unit experiences strictly less than $k$ disadvantages. Intuitively, $M$ captures the *mass of multiple disadvantages* by counting the MDSs of those with $k$ or more disadvantages (i.e., $\sum_{i=1}^{n} c_i \times \mathbb{I}[c_i \geq k]$), normalized by the maximum feasible number of disadvantages (i.e., $n \times d$). Alternatively, $M$ may be interpreted as an average of the censored normalized MDSs, i.e., $(c_i \times \mathbb{I}[c_i \geq k])/d$. Like $H$, $M$ is bounded between 0 and 1. When the disadvantage threshold is $k = d$, the identity $M(d) = H(d)$ holds.

Our primary outcome measure for multidimensional impact evaluation is $M$, but we also analyze the changes in $H$ and $A$ to examine how the overall change in $M$ is accomplished. Studying this breakdown has useful *policy implications*. A program that eliminates disadvantages among those experiencing lower MDSs will show a reduction in $M$ that is mainly driven by a reduction in $H$. On the contrary, if the program primarily eliminates disadvantages among those with high MDSs but do not necessarily bring their MDSs below $k$, then the reduction in $M$ will be driven by a reduction in $A$. To facilitate our understanding, recall our illustration involving $X, \overline{X}_1$ and $\overline{X}_2$. Suppose, $k = 2$. The pre-program mass, incidence, and intensity for $X$ are 5/12, 1/2, and 5/6, respectively. For $\overline{X}_1$, they are 1/6, 1/4, and 2/3, respectively. Thus, a 60 percent reduction in $M$ is accompanied by a 50 percent reduction in $H$ and a 20 percent reduction in $A$. Let us now look at $\overline{X}_2$, where the mass, incidence, and intensity are 1/4, 1/4, and 1, respectively. Unfortunately, in this case, the 40 percent reduction in $M$ is accompanied by a 50 percent reduction in $H$, *but* a 20 percent increase in $A$.

Our illustration above uses a particular disadvantage threshold $k$. The use of a range of thresholds, however, is helpful when evaluating a program's impact on the distribution of multiple disadvantages. The concept is analogous to poverty dominance (Atkinson, 1987; Foster and Shorrocks, 1988; Ravallion, 1994; Alkire et al., 2015). Let us go back to $X, \overline{X}_1$, and $\overline{X}_2$. If we consider the deprivation threshold to be $k = 1$, then three units in $X$ experience one or more disadvantages and the pre-program mass is 1/2. In $\overline{X}_1$, two units experience one or more disadvantages and the associated post-program mass is 1/4. The post-program mass in $\overline{X}_2$ is also 1/4. In both cases, the program has reduced the masses by 50 percent. However, for $k = 3$, the program exhibits a positive impact when $\overline{X}_1$ is obtained from $X$, but does not show any change when $\overline{X}_2$ is obtained from $X$. Therefore, with $\overline{X}_2$, the program cannot be considered inclusive because the unit with the greatest need did not benefit from the program's overall impact.

### 3. The Program, Data, and Econometric Analysis

The Philippine CCT program (4Ps) is the government's flagship poverty reduction strategy and human capital investment program. The program's primary objectives are to: (a) improve preventive health care among pregnant women and young children; (b) raise school enrollment and attendance rates among children; (c) reduce the incidence of child labor; and (d) raise the average food consumption expenditure of poor households (DSWD, 2012). Cash grants, which are provided upon fulfilling the required conditionalities, are expected to assist beneficiary households to achieve these objectives.

Program components described here are those applicable for the period covered in this evaluation study—2008–2011 (Fernandez and Olfindo, 2011). Beneficiary households get two types of cash grants, released every 2 months: the education grant and the health grant. The education grant is ₱ 300 per month or ₱ 3000 per year for each school-age children of 14 years or younger, for a maximum of three beneficiary children per household. At the time of data collection, the exchange rate is approximately US$1 = ₱ 45. The education grant is expected to cover schooling expenses and to compensate families for possible income losses because of the schooling conditionality.

The health grant is 500 per month or 6000 per year. All beneficiary households are entitled to this grant, which aims to improve food consumption. The maximum overall grant for each household is thus 15,000 per year. This amount is around 15 percent of the income poverty line when 4Ps was initiated in treatment areas. Based on the 4Ps grants data, the average annual grants received by treatment households between January 2009 and November 2011 is 9022. Actual grants received depend on household composition and compliance to program conditionalities listed in Table 1. The 4Ps followed a phased-in implementation design. Areas with the highest incidence of poverty based on 2006 poverty statistics were prioritized in 2008. By 2010, 4Ps was initiated in all provinces. *Beneficiary households* are identified as follows. First, a household is identified as *poor* if its predicted income, estimated through proxy means test (PMT), falls below the required poverty threshold.[5] Then, a poor household is identified as *eligible* if it has either at least one child aged 0–14 years or a pregnant member. Finally, *eligible households* are invited to a village assembly for information validation and to formalize program enlistment. In sum, *beneficiary households*: (1) reside in areas selected for the program; (2) are identified as poor through PMT; (3) have either children aged 0–14 years or a pregnant member; and (4) are validated eligible during a village assembly.

The Philippine government considers the 4Ps to be a major contributor to recent poverty reduction. It is claimed to have increased the average income among the bottom three deciles of the population by 82 percent and have reduced monetary poverty incidence from 26.3 percent to 21.6 percent between 2009 and 2015 (NEDA, 2017). Program evaluation reports show that the program improved outcomes

---

[5]Because of inherent difficulties in collecting income data directly, a PMT uses multivariate regression techniques to estimate incomes using correlate proxy indicators. The first wave of household listing, conducted between 2007 and 2011, identified 5.25 million of 10.9 million households as poor (see Fernandez, 2012).

TABLE 1
4Ps Conditionalities for Receiving Education and Health Grants

| Grant Type | Applicable Member | Conditionalities |
|---|---|---|
| Education | Children 3–14 years old | Enroll in school and maintain a class attendance rate of at least 85% per month |
| Health | Children 0–14 years old | For children 0–5 years old: <br>• Complete all required vaccinations <br>• Monthly weight monitoring and nutrition counseling for children aged 0–23 months old; bi-monthly for 24–72 months old <br>For children 6–14 years old in elementary level: <br>• Complete all required vaccinations <br>• Monthly weight monitoring and nutrition counseling for children aged 0–23 months old; bi-monthly for 24–72 months old |
| | Pregnant women | Need to satisfy all the following criteria: <br>• One pre-natal consultation each trimester <br>• At least one blood pressure and weight monitoring measurement in each trimester <br>• At least one Breastfeeding Counseling Session prior to delivery and within the first six weeks after childbirth <br>• At least one Family Planning Counseling Session prior to delivery and within the first six weeks after childbirth <br>• Delivery by a skilled health professional <br>• At least one post-natal care within the first six weeks after childbirth |
| | Mother or other designated guardian | Attend monthly Family Development Sessions (lectures on nutrition, sanitation, reproductive health, responsible parenthood, among others) |

Source: Department of Social Welfare and Development (2012). Pantawid Pamilya Operations Manual.

and reduced noncompliance rates in different indicators. Onishi, Friedman and Chaudhury (2013a), for instance, find positive impacts on school enrollment of 3- to 11-year-old children and on nutritional status of 6- to 36-month-old children. Similarly, Onishi, Kandpal, Friedman and Chaudhury (2013b) detect increases in the consumption of food and non-food items. Meanwhile, Orbeta, Abdon, del Mundo, Tutor, Valera and Yarcia (2014) observe improvements in school enrollment among 12- to 15-year-old children, in deliveries in health facilities, and in spending for education.

### 3.1. *Data and Experimental Design*

Admirably, the 4Ps is one of the few nationwide programs with an embedded impact evaluation design. Since 2011, three waves of impact evaluation surveys have been conducted to evaluate the program's causal impacts on health, education, and poverty outcomes. Each wave collects samples for both *randomized*

*control trial* (RCT) and *regression discontinuity design* (RDD) evaluation. In this paper, we use only the first wave of the RCT evaluation survey. We prefer an RCT survey over an RDD survey because the latter only captures localized treatment effect and may miss the program's impact on the households in greatest need of intervention. Meanwhile, since the second wave of the RCT survey, control households have been incorporated into the program.

The RCT survey follows a cluster randomized trial design, where treatment assignment is determined at the village level. In October 2008, eight municipalities are chosen to represent the poorest municipalities in the poorest provinces, and 130 *clusters* or *villages* are randomly drawn from these municipalities. Half of these villages are assigned to treatment. Program implementation in treatment areas commenced in January 2009, and the first wave of impact evaluation survey is carried out between October and November 2011. A total of 1418 sample households are surveyed—704 from treatment and 714 from control villages.[6] Treatment assignment is credibly implemented as there are no households from the control villages that received 4Ps benefits based on the beneficiary database.[7]Meanwhile, 8 percent of sample households in treatment areas are not 4Ps beneficiaries. Possibly, these households did not participate in the community assembly, where eligible beneficiaries confirm their information and register for the program. Alternatively, they may have opted out of the program or were dropped from the list of eligible households during community validation as inclusion errors (Onishi et al., 2013a).

Ideally, all sample households in the RCT survey should have PMT incomes below the respective provincial poverty thresholds and should have at least one program-eligible member. We observe, however, that around 9 percent of the sample households in the survey do not have any program-eligible member, potentially because of changes in household composition between the time of the household assessment in 2008 and the time of the first wave survey in 2011. There is no systematic difference between the treatment and control groups on the proportion of households without a program-eligible member. These households are dropped from the analysis.

### 3.2. *Econometric Specification and Experimental Validity*

Actual program status may be affected by realities on the ground, such as self-selection and other program implementation challenges. For our analysis, the eligible households residing in treatment villages are considered as *treated* regardless of actual program status and the eligible households in control villages are considered as *controls*. In the literature, this approach is referred to as estimating *intent-to-treatment* (ITT) effect, or the average potential impact of offering the program, which captures the change in outcomes among the eligible households *given the opportunity to participate*.

Our unit of analysis is the household and we estimate the causal impact of the 4Ps using the following regression specification:

[6]A power analysis conducted in Onishi, Friedman and Chaudhury (2013a) on three outcome variables—monthly per capita household consumption, school participation on 6- to 14-year-old children, and health facility visits of 0- to 5-year-old children—shows that a sample of 1300 is sufficient to detect impact with 80 percent power.

$$(4) \qquad y_i = \alpha + \tau p_i + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i,$$

where $y_i$ is the outcome variable for household $i$, $p_i$ is the binary program assignment such that $p_i = 1$ if household $i$ resides in treatment areas and $p_i = 0$ otherwise, $\tau$ estimates the program's ITT effect, $\mathbf{x}_i$ is a vector of covariates that we control and $\epsilon_i$ is the error term. A negative estimated value of $\tau$ reflects an improvement in the outcome variable and vice versa.

For estimating the impact on the intensity ($A$) or the average MDS of those experiencing $k$ or more disadvantages, we use the following regression specification:

$$(5) \qquad y_i = \alpha + \theta_1 z_i + \theta_2 p_i + \theta_3 \left(z_i \times p_i\right) + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i,$$

where $z_i$ is a binary variable taking a value of 1 whenever the MDS is larger than $k$ and a value of 0 otherwise and $p_i$ is the binary program assignment. Please note that we add an interaction term combining $z_i$ and $p_i$. The program's impact on the intensity is estimated from Equation 5 as:

$$(6) \qquad \tau = E(y \,|\, z = 1, p = 1) - E(y \,|\, z = 1, p = 0) = \theta_2 + \theta_3.$$

We conduct balance tests to check the credibility of the randomization and indeed we do not find any significant differences in the baseline characteristics between treatment and control groups (results are available in Supplementary Tables A1 and A2. As a full baseline survey is not available, we can only test demographics and household characteristics that are used to compute the PMT incomes. We find no reason to doubt that potential outcomes are not independent of treatment assignment. The PMT formula, used for identifying the poor, is not released to the public. The program's poverty thresholds are also set by the national statistics agency and not by the program implementer.

## 4. Effectiveness of the 4Ps in Inducing Behavioral Changes

In our first exercise, we examine whether the 4Ps has induced behavioral changes or reduced noncompliances vis-a-vis the program conditionalities. Several studies have examined the impact of the 4Ps on noncompliance rates for different indicators separately, but none has looked at the program's impact on joint noncompliances.

### 4.1. *Noncompliance Indicators and Sample Selection*

The indicators are drawn from the program's conditionalities listed in Table 1. Ideally, we ought to incorporate all conditionalities in our analysis, but our selection of indicators is constrained by our study objective as well as by the availability of data. For instance, we cannot include the information on family development sessions because the information is only applicable to households assigned to treatment. Similarly, we cannot use the information on certain conditionalities for pregnant women—such as monitoring sessions for blood pressure and weight and counseling sessions for breastfeeding and family planning—because these were not

TABLE 2

INDICATORS AND NONCOMPLIANCE CRITERIA FOR STUDYING THE 4PS' IMPACT ON MULTIPLE NONCOMPLIANCES

| Indicator | Noncompliance Criterion (Household Level) |
|---|---|
| Attendance | Household has at least one 3- to 14-year-old child with attendance rate below 85 percent |
| Health visit | Household has at least one 0- to 5-year-old child who did not have regular growth and nutrition monitoring visits |
| Deworming | Household has at least one 6- to 14-year-old child in elementary who did not receive two deworming pills |
| Prenatal visit | Household has at least one woman (currently pregnant or had live birth in the past 2 years) not having prescribed number of prenatal visit |
| Birth delivery | Household has any live birth in the past 2 years, but the birth is either not delivered in a health facility or by a health professional |

monitored separately for compliance (only visits to the health center were monitored). Finally, we are unable to include immunization and postnatal care because of missing data issues.[7]

Another challenge that we encounter among the remaining conditionalities is that the *applicable populations* are not uniform. For example, the education conditionalities are applicable to 3- to 14-year-old children, whereas the health conditionalities are applicable to 0- to 14-year-old children as well as to pregnant women. We included at least one indicator from each of the relevant target populations: school-age children (3–14 years old), 0- to 5-year-old children, and women of reproductive age. Given that each conditionality has respective *applicable population*, it is understandably not feasible for all households in our sample to have program-eligible member(s) for every conditionality.

We can select only five indicators that are directly targeted by the program. Table 2 presents the indicators and the noncompliance criteria. The *applicable populations* for the first three indicators are children of different age groups, whereas the *applicable population* for the final two indicators is women of reproductive age (i.e. 15–49 years old). Information on prenatal visit or birth delivery is available for female household members who are currently pregnant and have given a live birth in the past 5 years. However, considering the program's exposure from January 2009 to September 2011, we only consider births that are delivered from October 2009 onward. Therefore, to be included in the analysis, a child's growth must have been potentially "covered" by the program from the time of conception—a critical period in the child's development (UNICEF, 2014).

The challenge of nonoverlapping *applicable populations* also entails a crucial trade-off. Note that our multidimensional evaluation exercise requires us to look at households' noncompliance profiles across indicators jointly. Yet, only 25 percent of sample households have at least one program-eligible member for the *birth delivery* indicator, whereas 98 percent have at least one member for the *attendance*

---

[7]Vaccination details are particularly challenging to recall and are taken from immunization cards, which are presented for only around 24 percent of 0- to 5-year-old children. The postnatal care indicator suffers from similar significant missing data issues.

TABLE 3
PROPORTION OF HOUSEHOLDS WITH NONCOMPLIANCES BY SAMPLE TYPES

| | Sample A | | Sample B | | Sample C | | B − C |
|---|---|---|---|---|---|---|---|
| Panel I: Incidence of non-compliance per indicator | | | | | | | |
| Attendance | 1,266 | 0.398 | 1,023 | 0.377 | 243 | 0.486 | −0.108*** |
| Health visit | 752 | 0.763 | 509 | 0.747 | 243 | 0.798 | −0.052 |
| Deworming | 1,104 | 0.645 | 861 | 0.639 | 243 | 0.667 | −0.028 |
| Prenatal | 394 | 0.393 | 151 | 0.444 | 243 | 0.362 | 0.082 |
| Birth delivery | 322 | 0.677 | 79 | 0.646 | 243 | 0.687 | −0.042 |
| Panel II: Distribution of households by their number of non-compliances | | | | | | | |
| No non-compliance | 1,290 | 0.178 | 1,047 | 0.215 | 243 | 0.021 | 0.194*** |
| One non-compliance | 1,290 | 0.316 | 1,047 | 0.370 | 243 | 0.082 | 0.287*** |
| Two non-compliances | 1,290 | 0.264 | 1,047 | 0.266 | 243 | 0.251 | 0.015 |
| Three non-compliances | 1,290 | 0.160 | 1,047 | 0.129 | 243 | 0.296 | −0.167*** |
| Four non-compliances | 1,290 | 0.058 | 1,047 | 0.020 | 243 | 0.222 | −0.202*** |
| Five non-compliances | 1,290 | 0.024 | 1,047 | 0.000 | 243 | 0.128 | −0.128*** |

*Note*: Under column headings Sample A, Sample B, and Sample C, the left sub-column reports the numbers of sample households and the right sub-column reports the proportions of households. The final column (B–C) reports the differences of proportions between Sample B and Sample C and their statistical significance. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

indicator (the distribution of the number of eligible members per indicator among the 1290 sample households may be found in Supplementary Table A3).

Therefore, we may either restrict our attention to sample households with eligible members for all five selected indicators or we may consider all sample households with eligible member(s) in at least one indicator. The former option leads to a sample of merely 243 households, which severely reduces the statistical power and representativeness of our analysis. The choice of indicators with nonoverlapping *applicable populations* follows naturally from the design of the 4Ps intervention and is thus unavoidable.

To elucidate the loss of representativeness, we divide the sample containing all households with eligible member(s) in at least one indicator (Sample A) into a sample of households with eligible member(s) in at least one but fewer than five indicators (Sample B) and a sample of households with eligible members in all five indicators (Sample C). In Panel I of Table 3, we present the incidences of non-compliances of all five indicators within each sample, where the only statistically significant difference between Sample B and Sample C is observed for the *attendance* indicator. In Panel II, we present the distribution of households experiencing different numbers of noncompliances within each sample, where the distribution for Sample C is vastly different from the distribution of Sample B and thus from Sample A. Thus, we conduct our primary analysis on the entire sample of 1290 households (Sample A), but we also verify the robustness of our findings for the sample of 243 households (Sample C).

Considering the entire sample of 1290 households allows us to capture the impact of 4Ps without losing representativeness, but implicitly treats a household without any eligible member in an indicator to be compliant in that indicator. This approach is common for cross-country and inter-temporal comparisons in multidimensional poverty analysis (See, United Nations Development Programme, 2010; Alkire and Santos, 2014; Alkire et al., 2017). It is infeasible for every household,

TABLE 4

PROPORTION OF HOUSEHOLDS WITH NONCOMPLIANCES PER INDICATOR AND THE JOINT DISTRIBUTION OF
NONCOMPLIANCES ACROSS TREATMENT AND CONTROL GROUPS

| | Overall | | Control | | Treatment | | Difference |
|---|---|---|---|---|---|---|---|
| Panel I: Incidence of non-compliance per indicator | | | | | | | |
| Attendance | 1,266 | 0.398 | 635 | 0.454 | 631 | 0.342 | −0.111*** |
| Health visit | 752 | 0.763 | 371 | 0.833 | 381 | 0.696 | −0.137*** |
| Deworming | 1,104 | 0.645 | 553 | 0.685 | 551 | 0.604 | −0.081** |
| Prenatal | 394 | 0.393 | 195 | 0.431 | 199 | 0.357 | −0.074 |
| Birth delivery | 322 | 0.677 | 157 | 0.682 | 165 | 0.673 | −0.009 |
| Panel II: Distribution of joint non-compliances | | | | | | | |
| $M(k=1)$ | 1,290 | 0.335 | 649 | 0.360 | 641 | 0.311 | −0.049*** |
| $M(k=2)$ | 1,290 | 0.272 | 649 | 0.300 | 641 | 0.245 | −0.055*** |
| $M(k=3)$ | 1,290 | 0.167 | 649 | 0.191 | 641 | 0.142 | −0.049** |
| $M(k=4)$ | 1,290 | 0.071 | 649 | 0.075 | 641 | 0.066 | −0.008 |
| $M(k=5)$ | 1,290 | 0.024 | 649 | 0.022 | 641 | 0.027 | 0.005 |

*Note*: Under column headings Overall, Control, and Treatment, the left sub-column reports the numbers of sample households and the right sub-column reports the proportions of households. The final column (Difference) reports the unconditional differences in proportions between treatment and control households and their statistical significance. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

under this option, to be noncompliant in all five indicators, which may be crucial when targeting households by affecting inter-household comparability. As we do not conduct any targeting exercise, such comparability is not a concern for our analysis.

Given that a larger number of program-eligible members may make a household more likely to experience noncompliances in a larger number of indicators, we control for the number of 4Ps-eligible members at baseline when estimating the program's impact on each outcome.

In Table 4, considering a noncompliance as a disadvantage, we present descriptive statistics on unconditional noncompliance rates for all five indicators (Panel I) as well as the *masses of multiple noncompliances* (Panel II) for the overall sample, the treatment sample, and the control sample. The sample of treatment households has statistically lower unconditional noncompliance rates for three indicators—*attendance*, *health visit*, and *deworming*. Unconditional masses of multiple noncompliances also appear to be significantly lower among the treated households for the noncompliance thresholds of $k = 1, 2, 3$. The cross-tabulation of noncompliances across five indicators for the overall, treated, and control samples is available in Supplementary Table A4.

## 4.2. *Impact on Multiple Noncompliances*

We now estimate the program's impact on the *masses of multiple noncompliances* for different noncompliance thresholds ($k$). Our main outcome variable is the censored normalized multiple noncompliance score, i.e. $y_i = c_i/d$ if $c_i \geq k$ and $y_i = 0$ if $c_i < k$ in terms of the notation in Equation 4. We also estimate the program's impact on the incidence of noncompliances for different indicators as well as the incidence of multiple noncompliances for different $k$ values.

TABLE 5

ESTIMATES OF 4PS' IMPACT ON NONCOMPLIANCES FOR HOUSEHOLDS WITH ELIGIBLE MEMBER(S) IN AT LEAST ONE INDICATOR

| | Attendance | Health visit | Deworming | Prenatal | Birth delivery |
|---|---|---|---|---|---|
| | -0.115*** | -0.118*** | -0.078*** | -0.073 | -0.028 |
| | $[-0.163, -0.066]$ | $[-0.176, -0.061]$ | $[-0.125, -0.030]$ | $[-0.153, 0.007]$ | $[-0.120, 0.063]$ |
| | (0.454) | (0.833) | (0.685) | (0.431) | (0.682) |
| | $\langle 1,266 \rangle$ | $\langle 752 \rangle$ | $\langle 1,104 \rangle$ | $\langle 394 \rangle$ | $\langle 322 \rangle$ |
| Estimates of impact on the joint-distribution of non-compliances | | | | | |
| | $k = 1$ | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
| $M$ | -0.051*** | -0.059*** | -0.052*** | -0.010 | 0.005 |
| | $[-0.073, -0.029]$ | $[-0.084, -0.034]$ | $[-0.078, -0.026]$ | $[-0.033, 0.013]$ | $[-0.009, 0.018]$ |
| | (0.360) | (0.300) | (0.191) | (0.075) | (0.022) |
| | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ |
| $H$ | -0.061*** | -0.102*** | -0.084*** | -0.013 | |
| | $[-0.097, -0.024]$ | $[-0.145, -0.059]$ | $[-0.120, -0.047]$ | $[-0.040, 0.013]$ | |
| | (0.854) | (0.553) | (0.282) | (0.088) | |
| | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | |
| $A$ | -0.170*** | -0.039 | 0.117 | 0.099 | |
| | $[-0.268, -0.073]$ | $[-0.150, 0.071]$ | $[-0.027, 0.261]$ | $[-0.038, 0.236]$ | |
| | (2.106) | (2.708) | (3.388) | (4.246) | |
| | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | $\langle 1,290 \rangle$ | |

*Note*: A block of four rows presents the results for each outcome. In each column and in the first row of each block, the impact estimate denotes the *intent-to-treat* effect. For each impact estimate, we show the the 90% confidence interval (square brackets), the control mean (parenthesis), and the number of observations (angular brackets). Baseline control variables include household head's age and completed years of education, and the number of program-eligible members. Additional village-level controls (from the impact evaluation but not baseline survey) are the numbers of grade- and high schools, number of doctors and midwives, and the presence of a health center in the village. All regressions control for municipality level fixed effects and standard errors are robust to clustering at the village level. $M$ and $H$ range between 0 and 1, but $A$ ranges between $k$ and 5. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

Each outcome variable for evaluating the impact on incidences is a binary variable, such that $y_i = 1$ if household $i$ experiences noncompliance (or experiences multiple noncompliances in the case of $H$) and $y_i = 0$ otherwise. Estimating impact on incidences using Equation 4 is equivalent to using linear probability models. We test the robustness of our findings by computing marginal effects through probabilistic models and we arrive at similar analytic conclusions.

The top-half of Table 5 presents the estimated impacts on the incidence of noncompliance for the five indicators, and the bottom-half of the table presents the estimates on the masses of multiple noncompliances ($M$) for different noncompliance thresholds ($k$). We additionally report the estimated impacts on the incidences ($H$) and intensities ($A$) of multiple noncompliances (using Equations 5 and 6).

A block of four rows in each column corresponds to an outcome. The first row within each block denotes the *causal impact estimate*, and the other three rows report the 90 percent confidence interval of the estimate (square brackets), the counterfactual mean (parentheses), and the corresponding sample size (angular brackets), respectively. The sample size for each indicator corresponds to the number of households with eligible members in that indicator. We control for household

characteristics, municipality-level fixed-effects, and village-level variables (supply-side factors) that may affect the variability of the outcomes.

The program significantly improved three child-related indicators—*attendance*, *health visit*, and *deworming*, of magnitudes 11.5, 11.8, and 7.3 percentage points, respectively. The *health visit* and *deworming* indicators are highly program-specific, and therefore positive impacts show that the conditionalities are effective in inducing household behavioral changes. We are unable, however, to statistically detect changes in the incidences of noncompliances for the *prenatal* and *birth delivery* indicators.

In the bottom-half of Table 5, we present the program's impact on the masses of multiple noncompliances. Overall, when we consider the households with one or more noncompliances ($k = 1$), we observe that the mass decreased statistically significantly by 0.051 points or by 14.2 percent. Decomposing the mass across the incidence and intensity of multiple noncompliances, we observe that the reduction in the mass is accompanied by a reduction in the incidence from 85.4 percent by 6.1 percentage-points. At the same time, the intensity of multiple noncompliances of those experiencing one or more noncompliances is lower by 0.17 points on average, which is equivalent to slightly less than one-fifth of an indicator.

A reduction in $M$ is certainly a positive finding, but to look at the effect on the distribution of multiple noncompliances, let us examine the changes in masses for other thresholds. Masses for $k = 2$ and $k = 3$ improved significantly by 0.059 and 0.052 points or by 19.7 percent and 27.2 percent, respectively. These reductions are accompanied by even larger magnitudes of decreases in corresponding incidences, where the proportions of households with two or more and three or more noncompliances are lower by 10.2 and 8.4 percentage-points, respectively. In contrast, the corresponding intensities or the average noncompliance did not fall. These contrasting findings may suggest that the reductions in masses for $k = 2$ and $k = 3$ are obtained by alleviating noncompliances among those with two or three noncompliances while leaving the compliance profiles of those experiencing a larger number of noncompliances unchanged.

Our conjecture is supported by the findings for $k = 4$ and $k = 5$. Even though the mass is lower by 13.3 percent for $k = 4$, this reduction is not statistically significant. The reduction in the corresponding incidence is also around 18 percent relative to the initial level, but the magnitude of the reduction in absolute term is less than a quarter compared to the reductions for $k = 1, 2$, and 3. A similar narrative unfolds for $k = 5$, where it is sufficient to interpret the change in the mass as $M = H$, but the number of sample households with five noncompliances is not sufficient for a meaningful evaluation.

We thus observe a partial positive impact of the 4Ps on multiple noncompliances. The program reduced noncompliances among households with three or fewer noncompliances by inducing desired behavioral changes through cash and conditionalities. The program, nevertheless, does not appear to improve the average condition of households experiencing more noncompliances.

We present two robustness exercises in Supplementary Tables A5 and A6. Looking at the impact estimates based on the 243 sample households with eligible members in all five indicators, we observe that the absolute reduction in the mass for $k = 4$ is less than one-third of the absolute reduction in the mass for $k = 3$ and

is around half of the absolute reductions in the masses for both $k = 1$ and $k = 2$ (Supplementary Table A5). Most estimates are, however, statistically insignificant because of low statistical power, which we verified through inverse power analysis (Andrews, 1989). We also estimate the *average treatment effect on the treated* using the program information on actual beneficiary status of the households. None of the eligible households in control areas received treatment, but around 6 percent of the sample treatment households did not participate in the program. Our *treatment-on-treated* estimates (Supplementary Table A6) are consistent with our *intent-to-treat estimates*.

## 5. Effectiveness of the 4Ps in Reducing Poverty

In our second exercise, we look at the program's contribution to reducing multidimensional poverty by assessing its impact on the incidences and joint distribution of selected deprivation indicators. Although there is a healthy debate surrounding the particular forms of multidimensional poverty measures (Ravallion, 2011; Alkire et al., 2011), the value of understanding the joint distribution of deprivations is considered crucial (Ferreira, 2011). As before, we use the counting framework and use both monetary and non-monetary indicators. Here, a deprivation is considered as a disadvantage. The mass of multiple deprivations is the adjusted head count ratio (Alkire and Foster, 2011), which is a product of the incidence and the intensity of multiple deprivations, divided by the number of indicators. To distinguish the notation in this section from that used in Section 4, we denote the mass and the incidence of multiple deprivations by $M'$ and $H'$, respectively, and the threshold or the *poverty cut-off* by $k'$.

### 5.1. *Deprivation Indicators*

We select five indicators, chiefly based on three considerations. First, the selected indicators are related to program objectives, but are not directly targeted by the 4Ps' conditionalities. Second, each indicator can reflect changes in deprivations over a relatively short period, i.e. between January 2009 and October/November 2011. Unfortunately, deprivations in many indicators—such as access to public services or adult education—are crucial, but they remain static over a short period. Third, to circumvent potential endogeneity issues, we avoid indicators that are used for constructing PMT incomes that, in turn, are used to determine program eligibility.

In Table 6, we list the selected indicators and their deprivation criteria, where a household is considered deprived in an indicator if it fails to meet a subsistence standard or *deprivation criterion*. The first indicator is *consumption*, which is aligned with the program's objective of raising the average food consumption through health grants. The indicator identifies a household as deprived if the household's total consumption expenditure is so low that it is not even sufficient to cover the minimum subsistence level of food expenditure. Around 47.5 percent of households in the control group are deprived in this indicator.

We complement the *consumption* indicator with two additional indicators—*hunger* and *nutrition*. The *hunger* indicator aims to capture the household's

TABLE 6
INDICATORS AND DEPRIVATION CRITERIA FOR STUDYING 4Ps' IMPACT ON DEPRIVATIONS

| Indicator | Deprivation Criterion (Household Level) |
|---|---|
| Consumption | Household's total consumption expenditure is lower than the food poverty line |
| Hunger | Household members had experienced hunger in more than one occasion in the past 3 months |
| Nutrition | Household has at least one 0- to 5-year-old child, whose weight-for-age is two standard deviations lower than the median child growth standards |
| Dropout | Household has at least one 3- to 14-year-old child, who is not attending school |
| Savings | Household does not have a savings account **or** any other financial instrument |

deprivation in the availability of food. Each household was asked about the number of occurrences in the past 3 months when they experienced hunger and did not have anything to eat. A household is deprived if it experienced such a situation in more than one occasion. We avoid considering "one occasion of hunger" as a reflection of potential deprivation because a single occurrence may be because of recall issues or other external shocks unrelated to deprivation.

The *nutrition* indicator captures direct health deprivation within the household through child undernourishment assessed using the World Health Organization's growth standards on weight-for-age.[8] We do not observe strong overlaps between these three indicators, which indicate that they are capturing different aspects of deprivation in this particular context. In addition, in Supplementary Table A7, we present the proportion of households deprived in each of the five indicators as well as the proportion of households deprived simultaneously in each pair of indicators.

The fourth indicator—*dropout*—may appear to be the same as the attendance indicator that we use in Section 4, but it is not directly targeted by program conditionalities. The program has a rather stricter criterion, which requires not only a child to be enrolled (the complement of dropout), but also at least 85 percent attendance rate. Moreover, a household receives education grants for only a maximum number of three children. Every household is not necessarily aware which of their children are targeted, and the program also does not prevent households from using the grant to enroll all their children.

The fifth indicator—*savings*—aims to reflect financial deprivation and identifies a household as deprived if it does not have a savings account or any other savings instrument, such as provident fund, life insurance, or pre-need insurance. Maintaining a savings account aims to help beneficiaries smoothen their

[8]An alternative well-known indicator is height-for-age or *stunting*, which effectively reflects cumulative effect of chronic under-nutrition not only since birth but also since conception. Therefore, we consider the weight-for-age indicator to be more suitable for studying the effect of the program within a shorter horizon.

TABLE 7

PROPORTION OF HOUSEHOLDS WITH DEPRIVATION PER INDICATOR AND THE JOINT DISTRIBUTION OF
DEPRIVATION ACROSS TREATMENT AND CONTROL GROUPS

| | Overall | | Control | | Treatment | | Difference |
|---|---|---|---|---|---|---|---|
| Panel I: Incidence of deprivation per indicator | | | | | | | |
| Consumption | 1,290 | 0.478 | 649 | 0.476 | 641 | 0.480 | 0.004 |
| Hunger | 1,290 | 0.172 | 649 | 0.185 | 641 | 0.159 | −0.026 |
| Underweight | 716 | 0.365 | 349 | 0.350 | 367 | 0.379 | 0.029 |
| Dropout | 1,266 | 0.241 | 635 | 0.285 | 631 | 0.197 | −0.089*** |
| Savings | 1,290 | 0.856 | 649 | 0.866 | 641 | 0.846 | −0.020 |
| Panel II: Distribution of joint deprivations | | | | | | | |
| $M'\,(k'=1)$ | 1,290 | 0.389 | 649 | 0.399 | 641 | 0.379 | −0.020 |
| $M'\,(k'=2)$ | 1,290 | 0.331 | 649 | 0.341 | 641 | 0.320 | −0.021 |
| $M'\,(k'=3)$ | 1,290 | 0.190 | 649 | 0.210 | 641 | 0.169 | −0.040* |
| $M'\,(k'=4)$ | 1,290 | 0.074 | 649 | 0.087 | 641 | 0.062 | −0.025 |
| $M'\,(k'=5)$ | 1,290 | 0.012 | 649 | 0.014 | 641 | 0.009 | −0.005 |

*Note*: Under column headings Overall, Control, and Treatment, the left sub-column reports the numbers of sample households and the right sub-column reports the proportions of households. The final column (Difference) reports the unconditional differences in proportions between treatment and control households and their statistical significance. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

consumption or to open opportunities for other financial or enterprise assistance, both of which can lead to welfare improvement.

In Table 7, we present the unconditional deprivation rates for all five indicators (Panel I) and the *masses of multiple deprivations* (Panel II) for the overall sample, the treatment sample, and the control sample. We observe that treatment households have statistically lower unconditional deprivation rate only for the *dropout* indicator. Similarly, the unconditional mass of multiple deprivations is significantly lower among the treated households only for the deprivation threshold of $k' = 3$.

### 5.2. *Impact on Multiple Deprivations*

We now examine the program's causal impact on *multidimensional poverty* using the same sample of 1290 households with at least one program-eligible member. We use the regression specification in Equation 4 to estimate the causal impact of the 4Ps and the primary outcomes of interest are *deprivation incidences* of the indicators in Table 6 and the masses of multiple deprivations for different poverty cut-offs. The counting framework allows incorporating welfare weights to evaluate changes in joint deprivations, but we do not have such information available set by the program. The set of controls includes selected household characteristics, municipality-level fixed-effects and village-level characteristics.

In the top-half of Table 8, we report impact estimates on deprivation incidences of the five indicators. Interpretations of the components in the table are the same as that in Table 5. The program's impact on deprivation incidences for four of the five indicators (consumption, hunger, underweight, and savings) is small and statistically insignificant. The only statistically significant positive impact is observed for the dropout indicator. The underweight indicator somehow reflects a negative impact. Our overall finding about consumption deprivation is consistent

TABLE 8

ESTIMATES OF 4Ps' IMPACT ON DEPRIVATIONS FOR HOUSEHOLDS WITH ELIGIBLE MEMBER(S) IN AT LEAST ONE INDICATOR

| | Consumption | Hunger | Underweight | Dropout | Savings |
|---|---|---|---|---|---|
| | −0.015 | −0.022 | 0.041 | −0.088*** | −0.029 |
| | [−0.056, 0.026] | [−0.063, 0.019] | [−0.021, 0.103] | [−0.132, −0.043] | [−0.059, 0.000] |
| | (0.476) | (0.185) | (0.350) | (0.285) | (0.866) |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨716⟩ | ⟨1,266⟩ | ⟨1,290⟩ |
| Estimates of impact on the joint-distribution of deprivations | | | | | |
| $M'$ | $k' = 1$ | $k' = 2$ | $k' = 3$ | $k' = 4$ | $k' = 5$ |
| | −0.024** | −0.026* | −0.041** | −0.024* | −0.004 |
| | [−0.044, −0.004] | [−0.052, −0.000] | [−0.071, −0.012] | [−0.046, −0.002] | [−0.012, 0.005] |
| | (0.399) | (0.341) | (0.210) | (0.087) | (0.014) |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ |
| $H'$ | −0.008 | −0.021 | −0.058** | −0.029* | |
| | [−0.028, 0.012] | [−0.063, 0.022] | [−0.100, −0.016] | [−0.056, −0.003] | |
| | (0.926) | (0.639) | (0.310) | (0.105) | |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | |
| $A'$ | −0.116* | −0.124** | −0.047 | −0.011 | |
| | [−0.219, −0.014] | [−0.214, −0.035] | [−0.141, 0.047] | [−0.101, 0.079] | |
| | (2.153) | (2.670) | (3.383) | (4.132) | |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | |

*Note:* A block of four rows presents the results for each outcome. In each column and in the first row within each block, the impact estimate denotes the *intent-to-treat* effect. For each impact estimate, we show the 90% confidence interval (square brackets), the control mean (parenthesis), and the number of observations (angular brackets). Baseline control variables include household head's age and completed years of education and household size. Additional village-level controls (from the impact evaluation but not baseline survey) are the numbers of grade- and high schools. All regressions control for municipality level fixed effects and standard errors are robust to clustering at the village level. $M'$ and $H'$ range between 0 and 1, but $A'$ ranges between $k'$ and 5. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

with that of Onishi et al. (2013a), who also did not observe program impacts on consumption for the same period. Changes in consumption may potentially take more time to manifest, which was observed in case of the Mexico's CCT program (Fiszbein and Schady, 2009).

Does 4Ps exhibit a positive impact on the joint deprivations? In the lower-half of Table 8, we present the program's estimated impact on the masses of multiple deprivations for five different poverty cut-offs: $k' = 1, ..., 5$. For $k' = 1$ and $k' = 2$, the statistically significant reductions in masses are 0.024 and 0.026 points, or 6 percent and 7.6 percent, respectively. In both cases, these reductions are driven by decreases in the corresponding intensities. A potential reason for observing no changes in the incidences may be the prevalent deprivation for the savings indicator. For $k' = 3$, however, the statistically significant reduction in $M'$ is 0.041 points (or 19.5 percent) and is accompanied by 5.8 percentage points reduction in the incidence. Even for $k' = 4$, both $M'$ and $H'$ decrease significantly by around 27 percent. Thus, the 4Ps improved the overall masses and distribution of multiple deprivations.

To check the robustness of our findings, we compute the *intent-to-treat* estimates for the sample of 243 households with eligible members in all five indicators (Supplementary Table A8) as well as the *treatment-on-treated* estimates using the information on actual beneficiary status (Supplementary Table A9). Our findings are consistent with these additional estimates.

## 6. Is the Reduction in Poverty Shared by the Poorest?

Our findings in Section 4 show that the 4Ps, during our study period, did not seem to induce behavioral changes among those with four to five noncompliances. Section 5 results, however, show that the 4Ps, for the same period, improved the distribution of multiple deprivations. These findings raise certain questions. Are the households with four to five noncompliances the poorest among the beneficiary households? Is the overall reduction in multiple deprivations shared by those with four to five noncompliances?

First, we explore whether the households with four to five noncompliances are *associated* with experiencing more deprivations, on average, than the rest of the households. For convenience, we denote the multiple noncompliance score of household $i$ by $c_i^* \in [0, 5]$ and define a binary variable $T_i^{45}$, such that $T_i^{45} = 1$ if $c_i^* \geq 4$ and $T_i^{45} = 0$ otherwise. We use the following linear regression specification to explore the association:

$$(7) \qquad y_i^0 = \alpha_0 + \delta_0 T_i^{45} + \mathbf{x}_i^0 \boldsymbol{\beta}_0 + \epsilon_i^0,$$

where $y_i^0$ is the outcome of interest for household $i$ and $\delta_0$ estimates the difference in the averages of the outcome variable between those with four to five noncompliances and the rest of the households. Given that we are interested in the difference between the two groups before the intervention, the estimates are based only among the control group sample.

TABLE 9

COMPARISON OF DEPRIVATIONS BETWEEN HOUSEHOLDS WITH FOUR TO FIVE NONCOMPLIANCES AND THE REST OF ELIGIBLE HOUSEHOLDS

|  | Consumption | Hunger | Underweight | Dropout | Savings |
|---|---|---|---|---|---|
| $\hat{\delta}_0$ | 0.223*** | 0.041 | 0.093 | 0.400*** | 0.061 |
|  | [0.143, 0.304] | [−0.058, 0.139] | [−0.002, 0.189] | [0.309, 0.490] | [−0.017, 0.138] |
|  | (0.458) | (0.177) | (0.329) | (0.246) | (0.861) |
|  | ⟨649⟩ | ⟨649⟩ | ⟨349⟩ | ⟨635⟩ | ⟨649⟩ |
| Comparison of the joint distribution of deprivations in terms of ($M'$) | | | | | |
|  | $k' = 1$ | $k' = 2$ | $k' = 3$ | $k' = 4$ | $k' = 5$ |
| $\hat{\delta}_0$ | 0.193*** | 0.236*** | 0.268*** | 0.197*** | 0.074** |
|  | [0.147, 0.238] | [0.184, 0.288] | [0.194, 0.343] | [0.128, 0.267] | [0.019, 0.130] |
|  | (0.380) | (0.318) | (0.182) | (0.066) | (0.007) |
|  | ⟨649⟩ | ⟨649⟩ | ⟨649⟩ | ⟨649⟩ | ⟨649⟩ |

*Notes*: Estimates are based on control households in Sample A. A block of four rows presents the results for each outcome. In each column and in the first row in each block, $\hat{\delta}_0$ denotes the difference in the averages of each outcome between the households with 4-5 non-compliances and the households with 0-3 non-compliances. For each estimate, we show the the 90% confidence interval (square brackets), the mean outcome of the households with 0-3 non-compliances (parenthesis), and the number of observations (angular brackets). Baseline control variables include household head's age and completed years of education and household size. Additional village-level controls (from the impact evaluation but not baseline survey) are the numbers of grade- and high schools. All regressions control for municipality level fixed effects and standard errors are robust to clustering at the village level. $M'$ ranges between 0 and 1. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

We report the estimated differences in outcomes, $\hat{\delta}_0$, in Table 9, where the interpretations of its components are the same as that in Table 5. In the top half of the table, the $\hat{\delta}_0$ values reflect the differences in deprivation incidences for the five selected indicators. The households experiencing four to five noncompliances appear to be more deprived in all five indicators, but the differences are larger and statistically significant for the consumption and dropout indicators. Although statistically insignificant, the deprivation incidences in the hunger and underweight indicators are around 23–28 percent higher for households experiencing four to five noncompliances.

Meanwhile, the bottom half of the table presents the differences in the masses of deprivations ($M'$) for all five poverty cut-offs, $k' = 1, ..., 5$. We observe significantly lower masses for the households experiencing three or less noncompliances. For $k' = 1, ..., 4$, the estimated differences are between 0.193 and 0.268 points. Even for $k' = 5$, the estimated difference is 0.074 points. Thus, households with four to five noncompliances are associated with experiencing more extensive multidimensional poverty, on average, than the rest of the beneficiary households.

Now, the question is whether these poorer households benefited from the overall positive impact, albeit of small magnitude, of the 4Ps. We may answer this question by strictly focusing on the households that experience four to five noncompliances and examine the impact of the 4Ps within this group. This exercise, however, is not straightforward because selection to this group is determined after program assignment and the impact estimates may be subject to *selection bias*. For instance, affiliation to this group may be affected by both program eligibility, such as whether a household has at least one child or one pregnant member, and various supply-side factors, such as the availability of schools and health care centers.

To attenuate such bias, we use the Heckman sample selection procedure (Heckman, 1979). Based on Equation 4, the multiple noncompliance scores may be estimated as:

$$(8) \qquad c_i^* = \gamma_1 + \tau_1 p_i + \mathbf{x}_i \boldsymbol{\gamma} + \varepsilon_i,$$

where $p_i$ is the binary program assignment and $\varepsilon_i$ is the error term. Note that $T_i^{45} = 1$ whenever $c_i^* = \gamma_1 + \tau_1 p_i + \mathbf{x}_i \boldsymbol{\gamma} + \varepsilon_i \geq 4$. Therefore, the relevant sample selection equation is defined as: $T_i^{45} = 1 \left[ \gamma_1 + \tau_1 p_i + \mathbf{x}_i \boldsymbol{\gamma} + \varepsilon_i \geq 4 \right]$ or,

$$(9) \qquad T_i^{45} = 1 \left[ \gamma_1' + \tau_1 p_i + \mathbf{x}_i \boldsymbol{\gamma} + \varepsilon_i \geq 0 \right],$$

where $\gamma_1' = \gamma_1 - 4$.

The program's impact on outcomes (deprivation incidences and masses of multiple deprivations) among households experiencing four to five noncompliances is estimated by the following regression specification:

$$(10) \qquad y_i^{45} = \alpha_1 + \delta_{45} p_i + \mathbf{x}_i^1 \boldsymbol{\beta}_1 + u_i,$$

where $y_i^{45}$ is the outcome variable for household $i$ such that $T_i^{45} = 1$, $\delta_{45}$ is the coefficient for the program assignment variable $p_i$, $\mathbf{x}_i^1$ is a vector of covariates, and $u_i$ is the error term. The error terms $\varepsilon_i$ in Equation 9 and $u_i$ in Equation 10 are assumed to follow a bivariate normal distribution with zero means, standard deviations $\sigma_\varepsilon$ and $\sigma_u$, and correlation $\rho$. If $\hat{\rho} = 0$, then there is no sample selection problem (Wooldridge, 2010, p. 805), and the impact may be independently estimated by Equation 10. However, if $\hat{\rho} \neq 0$, then there is a sample selection problem and the impact should be estimated jointly by Equations 9 and 10.

Because we use the program assignment variable $p_i$ in the sample selection Equation 9 and also in Equation 10, the program's impact on each outcome among those experiencing four to five noncompliances cannot simply be estimated by $\hat{\delta}_{45}$. Instead, it should be estimated by $\hat{\tau}_{45} = \hat{\delta}_{45} + h(\hat{\rho}, \hat{\sigma}_\varepsilon, \hat{\sigma}_u, \hat{\gamma}_1', \hat{\tau}_1, \hat{\boldsymbol{\gamma}})$, where $h$ is a function of the estimated parameters from both equations. For the functional form of $h(\cdot)$, refer to Hoffmann and Kassouf (2005, Eq. 7).

We present our findings in Table 10, where the top-half reports the estimated impact on the deprivation incidences of five indicators, and the bottom-half shows the estimated impact on the masses of multiple deprivations. A negative estimate indicates that the poorer households benefited from the program. In fact, if the estimates are larger in magnitudes compared to the corresponding estimates in Table 8, then the impact would appear to be relatively favorable to the poorest of the poor. Interpretations of the components in the table are the same as that in Table 5, and each curly bracket reports the $p$-value for the Wald test of the null hypothesis $\hat{\rho} = 0$ (a rejection confirms the existence of sample selection problem). Detailed results of both the selection regression and the outcome regression are available in Supplementary Tables A10 and A11.

The consumption and savings indicators improved among households experiencing four to five noncompliances. The 4Ps induced 19 and 7.8 percentage points reductions in the deprivation incidences for consumption and savings, respectively.

TABLE 10

ESTIMATES OF 4PS' IMPACT ON DEPRIVATIONS AMONG HOUSEHOLDS EXPERIENCING FOUR TO FIVE NONCOMPLIANCES, CONTROLLING FOR SAMPLE SELECTION

| | Consumption | Hunger | Underweight[†] | Dropout | Savings |
|---|---|---|---|---|---|
| $\hat{\tau}_{45}$ | −0.190** | 0.075 | −0.022 | 0.034 | −0.078** |
| | [−0.319, −0.061] | [−0.067, 0.216] | [−0.242, 0.197] | [−0.156, 0.224] | [−0.140, −0.016] |
| | {0.508} | {0.281} | {0.000} | {0.752} | {0.824} |
| | (0.667) | (0.263) | (0.456) | (0.684) | (0.912) |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨716⟩ | ⟨1,266⟩ | ⟨1,290⟩ |
| Impact on the joint-distribution of deprivations ($M'$) | | | | | |
| | $k' = 1$ | $k' = 2$ | $k' = 3$ | $k' = 4$ | $k' = 5$ |
| $\hat{\tau}_{45}$ | -0.037 | -0.056 | -0.024 | 0.041 | -0.048 |
| | [−0.108, 0.034] | [−0.144, 0.032] | [−0.130, 0.082] | [−0.086, 0.169] | [−0.115, 0.019] |
| | {0.007} | {0.053} | {0.060} | {0.000} | {0.244} |
| | (0.596) | (0.582) | (0.498) | (0.298) | (0.088) |
| | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ | ⟨1,290⟩ |

*Notes*: A block of five rows presents the results for each outcome. In each column and in the first row in each block, $\hat{\tau}_{45}$ denotes the program's impact among households with 4-5 non-compliances in terms of the marginal effect conditional on $c_i^* \geq 4$ computed by the maximum likelihood estimation (MLE) method. For each impact estimate, we show the 90% confidence interval (square brackets), the p-value for the Wald test for rejecting the null hypothesis of $\hat{\rho} = 0$, where a rejection confirms the existence of sample selection (curly brackets), the control mean (parentheses), and the number of observations (angular brackets). Control variables for the selection equation (9) include program assignment, the number of program-eligible members at baseline, whether a household has at least one 0–2-year-old member, whether a household has at least one 3–5-year-old member, whether a household has at least one pregnant member, household head's age and completed years of education, and village level characteristics from the impact evaluation survey (not baseline)—the number of grade- and high schools, number of doctors and midwives, and the presence of a health center in the village. Control variables for Equation 10 include the number of household members at baseline, whether a household has a pregnant member, household head's age and completed years of education, and village level variables—the numbers of grade- and high schools. All regressions control for municipality fixed effects and standard errors are robust to clustering at the village level. $M'$ range between 0 and 1. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations. [†] The MLE process had convergence problems and we use the two-step Heckman estimation process instead. The *p*-value in the curly bracket corresponds to the *t*-test for rejecting the null hypothesis that the estimated coefficient for the inverse Mills ratio is zero.

These two magnitudes are, in fact, substantially larger than the corresponding magnitudes of impact estimates in Table 8. Although the program did not induce behavioral changes among the poorer households enough to hurdle the conditionalities, it successfully improved their consumption and saving behavior.

The deprivation incidences in hunger and dropout appear to have deteriorated among the poorer households, albeit statistically insignificantly. The result on the dropout indicator, in particular, is unsatisfactory because the program induced 8.8 percentage points reduction in the overall incidence of deprivation (Table 8).

Finally, we examine the program's impact on the masses of multiple deprivations or multidimensional poverty. From Table 8 we already observe that the overall masses for $k' = 1, 2, 3, 4$ decreased statistically significantly by 0.024–0.041 points. Among the households with four to five noncompliances in Table 10, however, we only observe a marginally larger (0.037–0.056 points) but statistically insignificant estimated reduction in the corresponding masses for $k = 1, 2$. Therefore, we observe that the 4Ps is not sufficiently inclusive in terms of reducing poverty and deprivations among the poorest of the poor.

## 7. Concluding Remarks

Our key objective in this paper is to formalize a multidimensional evaluation framework and empirically show that a causal evaluation exercise that incorporates the joint distribution of outcome is superior to conventional evaluation exercises. When the program in consideration addresses multiple disadvantages, improvements in a number of targeted outcomes evaluated separately may not be enough to fulfill the United Nations' pledge of "leave no one behind."

Thus, the question we tried to address is: *Do anti-poverty programs reach beneficiaries that require assistance in multiple outcomes simultaneously*? From a *policy perspective*, if these beneficiaries appear to be left out by the program, an immediate action point is to find out the reasons and reformulate the program's components.

We use the well-known counting framework (Atkinson, 2003; Alkire and Foster, 2011) for evaluating changes in the distribution of multiple outcomes. We then apply the framework to evaluate the impact of the Philippine CCT program—called Pantawid Pamilyang Pilipino Program (4Ps)—on beneficiary households between 2009 and 2011, using the first wave of randomized control trial survey embedded within the program.

Our empirical investigation reveals that 4Ps cash grants induced targeted behavioral changes among beneficiaries with fewer noncompliances, but those with a larger number of simultaneous noncompliances appear to have been left out. We further find out that, on average, the beneficiaries with more noncompliances experience more extensive deprivations as well. Although 4Ps reduced their consumption and savings deprivation, we find no conclusive evidence that they benefited from the overall poverty reduction. Previous evaluation studies on the 4Ps may have found satisfactory results in improving targeted outcomes separately, but our results show that program enhancements are needed to fulfill the UN pledge of "leave no one behind."

Our findings suggest that for the poorest families, the cash grants may only be enough to marginally improve their consumption, but are not sufficient to alleviate other associated deprivations. This particular observation calls for the need to specifically examine how poorer families manage program compliances as well as the required complementary interventions to ensure that they are not left behind.

Our results also highlight that improvements in each noncompliance matter in the ability of households to reduce their masses of multiple noncompliances, suggesting that impacts may be amplified if program components specifically address each noncompliance.

Currently, the education grant is targeted so that each child receives a grant to induce school attendance. The health grant is broad in that all households get the same amount, but the conditionalities they must satisfy depend on household composition. At the minimum, the household grantee needs to attend a monthly community meeting to qualify for the health grant. If the household happens to have a pregnant member or a 0- to 5-year-old child, it needs to satisfy all the relevant conditionalities (pre- and postnatal check-ups, facility-based delivery, and child weight monitoring) on top of the monthly meeting for the same health grant. If the program can develop levers to induce compliance to each of the health

conditionalities, then a household with a larger number of simultaneous noncompliances may be better equipped to reduce their noncompliances, which can set them toward a path of reduced mass of noncompliances. One starting point thus is to consider adjusting the health grants based on household composition.

## References

Alkire, S., and J. Foster, "Counting and Multidimensional Poverty Measurement," *Journal of Public Economics*, 95, 476–87, 2011.

Alkire, S., J. E. Foster, and M. E. Santos, "Where Did Identification Go?" *Journal of Economic Inequality*, 9, 501–05, 2011.

Alkire, S., J. E. Foster, S. Seth, M. E. Santos, J. M. Roche, and P. Ballon, *Multidimensional Poverty Measurement and Analysis*. Oxford University Press, Oxford, 2015.

Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymour, and A. Vaz, "The Women's Empowerment in Agriculture Index," *World Development*, 52, 71–91, 2013.

Alkire, S., J. M. Roche, and A. Vaz, "Changes Over Time in Multidimensional Poverty: Methodology and Results for 34 Countries," *World Development*, 94, 232–49, 2017.

Alkire, S., and M. E. Santos, "Measuring Acute Poverty in the Developing World: Robustness and Scope of the Multidimensional Poverty Index," *World Development*, 59, 251–74, 2014.

———, "Acute Multidimensional Poverty: A New Index for Developing Countries," *OPHI Working Paper 38*. Oxford: University of Oxford, 2010.

Andrews, D. W. K., "Power in Econometric Applications," *Econometrica*, 57, 1059–90, 1989.

Atkinson, A., "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches," *Journal of Economic Inequality*, 1, 51–65, 2003.

Atkinson, A. B., "On The Measurement of Poverty," *Econometrica*, 55, 749–764, 1987.

Azevedo, V., and M. Robles, "Multidimensional Targeting: Identifying Beneficiaries of Conditional Cash Transfer Programs," *Social Indicators Research*, 112, 447–75, 2013.

Baird, S., C. McIntosh, and B. Özler, "Cash or Condition? Evidence from a Cash Transfer Experiment," *Quarterly Journal of Economics*, 126, 1709–1753, 2011.

Bossert, W., S. R. Chakravarty, and C. D'Ambrosio, "Multidimensional Poverty and Material Deprivation with Discrete Data," *Review of Income and Wealth*, 59, 29–43, 2013.

Chakravarty, S. R., and C. D'Ambrosio, "The Measurement of Social Exclusion," *Review of Income and Wealth*, 52, 377–98, 2006.

Council of Economic Advisers, "Expanding Work Requirements in Non-Cash Welfare Programs," *Executive Office of the President of the United States*, Washington DC, 2018. https://www.white house.gov/wp-content/uploads/2018/07/Expanding-Work-Requirements-in-Non-Cash-Welfare-Programs.pdf.

Das, J., Q.-T. Do, and B. Özler, "Reassessing Conditional Cash Transfer Programs," *World Bank Research Observer*, 20, 57–80, 2005.

Department of Social Welfare and Development. *Pantawid Pamilya Operations Manual*. Quezon City: Department of Social Welfare and Development, 2012.

Dutta, I., and A. Mishra, "Measuring Vulnerability Using the Counting Approach," *Economics Discussion Paper Series EDP-1810*. Manchester: University of Manchester, Department of Economics, 2018.

Dutta, P., R. Murgai, M. Ravallion, and D. van de Walle, "Does India's Employment Guarantee Scheme Guarantee Employment?" *Economic and Political Weekly*, 47, 55–64, 2014.

Evans, D. K., and A. Popova, "Cash Transfers and Temptation Goods," *Economic Development and Cultural Change*, 65, 189–221, 2017.

Fernandez, L., "Design and Implementation Features of the National Household Targeting System in the Philippines," *Philippine Social Protection Note 5*, World Bank and Australian Aid, 2012). http://documents1.worldbank.org/curated/en/687351468093838577/pdf/701490BRI0P1180l0Pro tection0Note005.pdf.

Fernandez, L., and R. Olfindo, "Overview of the Philippines' Conditional Cash Transfer Program: The Pantawid Pamilyang Pilipino Program (Pantawid Pamilya)," *Social Protection Discussion Papers and Notes 69422*, The World Bank, 2011.

Ferreira, F. H. G., "Poverty is Multidimensional. But What Are We Going to Do About It?," *Journal of Economic Inequality*, 9, 493–95, 2011.

Filmer, D., and N. Schady, "Does More Cash in Conditional Cash Transfer Programs Always Lead to Larger Impacts on School Attendance?" *Journal of Development Economics*, 96, 150–57, 2011.

Fiszbein, A., and N. Schady, *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington DC: The World Bank, 2009.

Foster, J. E., "A Class of Chronic Poverty Measures," in T. Addison, D. Hulme, and R. Kanbur (eds), *Poverty Dynamics: Interdisciplinary Perspectives, 59–76*, Oxford University Press, Oxford, 2009.

Foster, J. E., and A. F. Shorrocks, "Poverty Orderings," *Econometrica*, 56, 173–7, 1988.

García, S., and J. E. Saavedra, "Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis," *Review of Educational Research*, 87, 921–965, 2017.

Glassman, A., Duran, D., Fleisher, L., Singer, D., Sturke, R., Angeles, G., Charles, J., Emrey, B., Gleason, J., Mwebsa, W., Saldana, K., Yarrow, K. & Koblinsky, M. "Impact of Conditional Cash Transfers on Maternal and Newborn Health," *Journal of Health, Population, and Nutrition*, 31(4 Suppl 2), S48–66, 2013.

Heckman, J., "Sample Selection Bias As A Specification Error," *Econometrica*, 47, 153–161, 1979.

Hoffmann, R., and A. L. Kassouf, "Deriving Conditional and Unconditional Marginal Effects in Log Earnings Equations Estimated by Heckman's Procedure," *Applied Economics*, 37, 1303–11, 2005.

deJanvry, A., and E. Sadoulet, "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality," *World Bank Economic Review*, 20, 1–29, 2006.

Loschmann, C., C. R. Parsons, and M. Siegel, "Does Shelter Assistance Reduce Poverty in Afghanistan?" *World Development*, 74, 305–322, 2015.

National Economic and Development Authority, *Philippine Development Plan 2017–2022*, National Economic and Development Authority,Philippines, 2017.

OECD, "Combating Poverty and Social Exclusion," *OECD Economic Surveys: France*, OECD Publishing, Paris, 2007. https://doi.org/10.1787/eco_surveys-fra-2007-4-en.

———, *A Broken Social Elevator? How to Promote Social Mobility*, OECD Publishing, Paris, 2018. https://doi.org/10.1787/9789264301085-en.

Onishi, J., J. Friedman, and N. Chaudhury, *Philippines Conditional Cash Transfer Program Impact Evaluation 2012 (English)*, World Bank Group,Washington, DC, 2013a.

Onishi, J., E. Kandpal, J. Friedman, and N. Chaudhury, *Evaluating Pantawid Pamilya using Regression Discontinuity Design: Key Results and Lessons (78134-PH)*, Social Protection Unit, East Asia and Pacific Region, The World Bank, 2013b.

Orbeta, A., A. Abdon, M. del Mundo, M. Tutor, M. T. Valera, and D. Yarcia, *Keeping Children Healthy and In School: Evaluating the Pantawid Pamilya using Regression Discontinuity Design, Second Wave Impact Evaluation Results*, Department of Social Welfare and Development, Quezon City, 2014.

Pasha, A., "Impact of Cash Grants on Multidimensional Poverty in South Africa," *Discussion Paper 208*, Courant Research Centre, Georg-August-Universität Göttingen, 2016.

Ravallion, M., "On Multidimensional Indices of Poverty," *Journal of Economic Inequality*, 9, 235–48, 2011.

———, *Poverty Comparisons*, Harwood Academic Press, Switzerland, 1994.

Robano, V., and S. C. Smith, "Multidimensional Targeting and Evaluation: A General Framework with an Application to a Poverty Program in Bangladesh," *Discussion Paper No. 7593*, Institute for the Study of Labor, Germany, 2013.

Song, S., and K. S. Imai, "Does the Hunger Safety Net Programme Reduce Multidimensional Poverty? Evidence from Kenya," *Development Studies Research*, 6, 47–61, 2019.

UNICEF, *The Right Ingredients: The Need to Invest in Child Nutrition*. London: UNICEF, 2014.

United Nations Development Programme, *The Real Wealth of Nations: Pathways to Human Development*, Palgrave Macmillan, New York, 2010.

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts, 2010.

World Bank, *Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle*, World Bank, Washington, DC, 2018.

## Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

**Table A1:** Balance tests of baseline household characteristics among households with eligible member(s) in at least one indicator

**Table A2:** Balance tests of baseline household characteristics among households with eligible members in all indicators

**Table A3:** Distribution of 4Ps-eligible members in our analysis sample households across treatment and control groups

**Table A4:** Proportion of households with non-compliance in each indicator and with simultaneous non-compliance in each pair of indicators

**Table A5:** 4Ps' impact estimates on non-compliances among households with eligible members in all indicators

**Table A6:** Treatment-on-the-treated estimates of 4Ps' impact on non-compliances among households with eligible member(s) in at least one indicator

**Table A7:** Proportion of households deprived in each indicator and simultaneously deprived in each pair of indicators

**Table A8:** 4Ps' impact estimates on deprivations among households with eligible members in all indicators

**Table A9:** Treatment-on-the-treated estimates of 4Ps' impact on deprivations among households with eligible members in all indicators

**Table A10:** The Heckman regression estimates for households with 4-5 non-compliances