Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford

OPHI *Working Paper No. 132*

Evaluation of Anti-poverty Programs' Impact on Joint Disadvantages: Insights from the Philippine Experience

Suman Seth* and Melba V. Tutor**

November 2019

Abstract
Anti-poverty programs increasingly target disadvantages in multiple outcomes to address current and future poverty. Conventional evaluation exercises, however, mostly estimate programs' impacts separately. We present a framework, drawing from the counting approach, that captures the joint distribution of disadvantages and allows the evaluation of programs' impacts on multiple disadvantages. We apply the framework to scrutinise the Philippine conditional cash transfer program using an embedded randomised control trial survey. Examining the program's impact on the distribution of multiple disadvantages, we observe that the program successfully reduced multiple disadvantages overall, but did not necessarily benefit the families experiencing a higher number of disadvantages simultaneously. Our results exemplify the valuable contribution of considering the joint distribution of disadvantages in evaluating anti-poverty programs' impacts.

Keywords: Impact evaluation, multidimensional poverty, joint distribution, conditional cash transfers, randomised control trial, Philippines

JEL Classification: C21, C51, I32, I38

*Economics Division, Leeds University Business School, University of Leeds, Leeds, LS2 9JT, UK, and Oxford Poverty & Human Development Initiative (OPHI), University of Oxford, UK, Email: S.Seth@leeds.ac.uk.
**Planning Service, Department of Agrarian Reform, Elliptical Road, Diliman, Quezon City, Philippines 1107, Email: mvtutor@upd.edu.ph.

This study has been prepared within the OPHI theme on multidimensional measurement.

Acknowledgments
We thank the participants of the OPHI Multidimensional Impact Evaluation workshop in March 2018 at the University of Oxford; the OPHI Lunchtime Seminar Series in February 2019 at the University of Oxford; the workshop on Recent Development in Distributional Analysis in April 2019 at the University of Leeds; the Faculty Colloquium Series Lecture in April 2019 at the Department of Economics, Presidency University, Kolkata; and the 3rd Development Economics Conference (DEC) in June 2019 at the University of Lincoln for their valuable comments. We also acknowledge the *Pantawid Pamilyang Pilipino* Program Management Office for sharing the data and for assistance during various stages of this study.

Citation: Seth, S. and Tutor, M. J. (2019): 'Evaluation of anti-poverty programs' impact on joint disadvantages: Insights from the Philippine experience' OPHI Working Paper 132, University of Oxford.

The Oxford Poverty and Human Development Initiative (OPHI) is a research centre within the Oxford Department of International Development, Queen Elizabeth House, at the University of Oxford. Led by Sabina Alkire, OPHI aspires to build and advance a more systematic methodological and economic framework for reducing multidimensional poverty, grounded in people's experiences and values.

Oxford Poverty & Human Development Initiative (OPHI)
Oxford Department of International Development
Queen Elizabeth House (QEH), University of Oxford
3 Mansfield Road, Oxford OX1 3TB, UK

Tel. +44 (0)1865 271915 Fax +44 (0)1865 281801
ophi@qeh.ox.ac.uk http://www.ophi.org.uk

ISSN 2040-8188 ISBN 978-19-1229-123-6

1 Introduction

Poverty alleviation strategies and anti-poverty programs are a fundamental component of welfare policies in both developed and developing countries. They range from a variety of welfare programs in the United States (e.g., see CEA, 2018) and strategies for tackling poverty, social exclusion, and social immobility in 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, and school-feeding programs targeted to poorer segments of the population.

Due to the multidimensional nature of poverty, anti-poverty programs implicitly target disadvantages in multiple outcomes simultaneously.[1] Anti-poverty programs targeted to the poorest segment of the population lend themselves naturally to multidimensionality since these people are more likely to experience simultaneous and multiple disadvantages. Moreover, program **theories of change** rest on addressing these disadvantages simultaneously to break the intergenerational cycle of poverty.[2] Yet, conventional program evaluation exercises mostly examine impacts separately. There is a need to look more closely at how anti-poverty programs affect desired outcomes jointly.

In this paper, we first present and justify a framework for assessing an anti-poverty program's impact from a multidimensional perspective based on the counting approach.[3] The counting framework is especially useful when the underlying indicators take binary forms, i.e., when each indicator is categorised into those who experience a disadvantage versus those who do not. At the same time, the framework enables capturing the joint distribution of disadvantages. We present how the framework allows evaluating a program's impact by analysing the changes in the **incidences** of people with multiple disadvantages as well as by examining the changes in the overall **masses** of disadvantages.

[1]In practice, the multidimensionality of poverty has been widely acknowledged by prominent international organisations. Within the United Nations, the UNDP (2010) has adopted the global Multidimensional Poverty Index (Alkire and Santos, 2010) and the multidimensionality of poverty has been embedded within the Sustainable Development Goals framework (https://sustainabledevelopment.un.org/sdg1); whereas, the World Bank (2018) has also attempted its first exercise in multidimensional global poverty measurement.

[2]For instance, cash grants, through conditional cash transfer programs, aim to tide-over families from chronic hunger (present poverty), while simultaneously incentivising access to schooling and healthcare to arrest future poverty (Fiszbein and Schady, 2009).

[3]For a discussion on the counting approach to poverty measurement and its contrast with the social welfare approach, see Atkinson (2003); for an axiomatic presentation, see Alkire and Foster (2011); and for applications of counting approaches, see Alkire et al. (2015, Chapter 4).

We then apply our framework to determine whether an anti-poverty program effectively addresses multiple disadvantages. In particular, we study a conditional cash transfer (CCT) program in Philippines, referred to as *Pantawid Pamilyang Pilipino Program* (4Ps).[4] Among various anti-poverty programs, CCTs have gained enormous popularity in recent decades as a key intervention mechanism for alleviating and breaking the intergenerational cycle of poverty. CCTs provide cash grants to beneficiary families conditional on compliance with pre-specified human capital investments. Their popularity stems from successes in various short-term outcomes: increased school attendance, fewer school drop-outs, lower discriminatory access to schooling for gender and minority groups, better access to child health care (immunisation, nutrient supplements, and health monitoring), maternal health care (prenatal care and facility-based deliveries), and higher food consumption expenditure.[5]

The cash grants in CCT programs aim to induce targeted behavioural changes among beneficiary households (Das et al., 2005). Therefore, a natural concern is whether these cash grants are reducing non-compliance in the targeted indicators. We investigate this by utilising a household survey specifically designed to capture the impact of 4Ps through randomised control trials. We select five indicators that are closely aligned with the 4Ps' conditionalities. We observe considerable reductions in the incidences of non-compliance (i.e., positive impact) in three indicators, confirming targeted behavioural changes in these indicators. When we examine changes in the joint distribution of multiple non-compliances using the counting framework, however, our analysis reveals unsatisfactory results. Although there is an overall reduction in joint non-compliance, we do not find significant improvements among families experiencing four or more non-compliances.

The chief objective of any anti-poverty program, nevertheless, is not just to reduce non-compliances but to improve welfare, and a CCT program is no exception. In the context of 4Ps then, we are concerned about (i) whether the families with four or more non-compliances are the poorest of the poor and (ii) whether 4Ps improved their living conditions. Hence, we want to find out whether a large and reportedly successful anti-poverty program is inclusive or pro-poorest. We examine this by selecting a set of five indicators that are not directly conditioned by the program but still capture different forms of deprivations.

---

[4] 4Ps has served 4.6 million beneficiaries and the Philippine government considers it to be a major contributor to recent poverty reduction. In fact, the President legally institutionalised the program through the "Pantawid Pamilyang Pilipino Program (4Ps) Act" on April 17, 2019.

[5] For discussions about CCTs' impact, see Fiszbein and Schady (2009), Filmer and Schady (2011), Baird et al. (2011), Glassman et al. (2013), Evans and Popova (2017), and García and Saavedra (2017). For critical evaluations, see Baird et al. (2011), Filmer and Schady (2011) and de Janvry and Sadoulet (2006).

We observe that families with four or more non-compliances experience more incidences of deprivation, on average, than the rest of the beneficiaries. We then implement the Heckman selection procedure to find out the program's impact on deprivations among these poorest households. Even though the program did not induce the poorest to make the behavioural changes necessary to comply with program conditionalities, we observe that it successfully improved their consumption. Apart from this, however, we neither observe reductions in incidences of deprivation for the other indicators, nor do we observe any conclusive pro-poorest improvement in the joint distribution of their deprivations. Our findings highlight the need for anti-poverty programs' impact evaluation exercises to not only examine disadvantages separately, but also their joint distribution.

The rest of our paper is organised as follows. We introduce the counting framework for evaluating a program's impact in Section 2. We present the overview of the Philippine CCT program in Section 3. We analyse the program's impact on non-compliance with the program conditionalities in Section 4 and examine whether the program has been inclusive in reducing deprivations among the poorest in Section 5. We provide concluding remarks in Section 6.

2 Impact evaluation from a multidimensional perspective

We use the term **disadvantage**, depending on the context, to refer either to non-compliance that reflects a failure to satisfy a condition pre-specified by a program, or to deprivation that reflects a failure to meet a minimum requirement of well-being. Let us illustrate how assessing impacts on different indicators separately precludes understanding whether the program benefited those who are disadvantaged in multiple indicators simultaneously.

Suppose an anti-poverty program directly targets three indicators. Let the following three matrices—X, $\bar{X}_1$ and $\bar{X}_2$—summarise the disadvantage profiles of four units, which may represent individuals or households. In each matrix, a row summarises the disadvantage profile of a unit in three indicators; whereas, a column summarises 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 the program's intervention. Otherwise, the unit does not experience any disadvantage ('ND') in the indicator.

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

Before the program, i.e., in X, two out of four units experience disadvantages in each of the three indicators. After the program, one of the two alternative disadvantage profiles, $\bar{X}_1$ and $\bar{X}_2$, may be obtained from X. The incidence or the proportion of units experiencing disadvantage within each indicator is now a quarter, both in $\bar{X}_1$ and $\bar{X}_2$. Thus, if the impact is evaluated for each indicator separately, then the program may appear to be equally effective, whether $\bar{X}_1$ or $\bar{X}_2$ 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, $\bar{X}_1$ is obtained from X by eliminating all three disadvantages of the fourth unit, while $\bar{X}_2$ 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 due to the program on average, but it leaves 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, $\bar{X}_1$ and $\bar{X}_2$, for instance, a unit is assigned a score of 1 for a status of 'D' and 0 for a status of 'ND'. All disadvantage status scores are summarised by an $n \times d$-dimensional disadvantage status score matrix $\mathcal{G}$, where a row contains the disadvantage profile of a unit.

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 $c_i = \sum_{j=1}^{d} g_{ij}$.[6] Clearly, c_i ranges between 0 and d for all i and a higher MDS reflects a

[6]One may consider different disadvantages of unequal importance and weight disadvantages unequally, which

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 who 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 normative judgement. For instance, if an evaluator aims 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.

2.1 Evaluating impact on multiple disadvantages

A straightforward evaluation exercise may be 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 in $\mathcal{G}$ for a given disadvantage threshold k by:

$$H(\mathcal{G};k) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}[c_i \geq k] = \frac{q_k}{n}; \tag{1}$$

where $\mathbb{1}[c_i \geq k]$ 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 and vice versa. Let us consider an example recalling the pre- and post-program disadvantage profiles X, $\bar{X}_1$ and $\bar{X}_2$. Suppose the program evaluator is interested in the program's impact on those experiencing two or more disadvantages (i.e., $k = 2$). As two units in X experience two or more disadvantages, the pre-program incidence is $1/2$. The associated post-program incidences for both $\bar{X}_1$ and $\bar{X}_2$ are $1/4$. In both cases, the program reduced the incidence of two or more disadvantages by 25 percentage points.

An impact evaluation exercise based only on comparing incidences, however, may ignore any change in the **intensity** or multiplicity of disadvantages among those who experience k or more disadvantages.[7] A simple way to reflect the intensity of multiple disadvantages may

is common in the multidimensional evaluation of well-being and poverty.

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

be to look at the average MDS of those experiencing k or more disadvantages:

$$A(\mathscr{G};k) = \frac{1}{q_k}\sum_{i=1}^{q_k}[c_i \mid c_i \geq k]. \tag{2}$$

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 who experience all d disadvantages (i.e., $k = d$) and one or more units have such experience.

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

Let us illustrate why an impact evaluation exercise should incorporate intensity in addition to the incidence of multiple disadvantages. Suppose, deprivation profile $\bar{X}_3$ 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, since two units in both X and $\bar{X}_3$ experience two or more disadvantages. Yet, the intensity of those experiencing two or more disadvantages (third and fourth units) declined from 2.5 in X to 2 in $\bar{X}_3$. 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.

Changes in both incidence and intensity of multiple disadvantages may be captured by the following measure, motivated by the **adjusted headcount ratio** (Alkire and Foster, 2011):[8]

$$M(\mathscr{G};k) = H(\mathscr{G};k) \times \frac{A(\mathscr{G};k)}{d} = \frac{1}{nd}\sum_{i=1}^{n} c_i \times \mathbb{1}[c_i \geq k]; \tag{3}$$

where $\mathbb{1}[c_i \geq k]$ is an indicator function. Measure M is a product of both incidence and intensity of multiple disadvantages divided by the number of indicators. Intuitively, M cap-

[8] The adjusted headcount ratio has numerous empirical applications. It is used to construct the well-known Multidimensional Poverty Index (UNDP, 2010; Alkire and Santos, 2014). Alkire and Seth (2015) and Alkire et al. (2017) used it to study changes in multidimensional poverty over time in India and in several developing countries, respectively. Loschmann et al. (2015) used it to study a shelter assistance program in Afghanistan; while, Pasha (2016) used it to examine the impact of social assistance grants in South Africa. Bag and Seth (2018) applied it to analyse the multidimensional standard of living within slums in India. For further applications, see Alkire et al. (2015, Chapter 5).

tures the **mass of multiple disadvantages** by counting the MDSs of those with k or more disadvantages (i.e., $\sum_{i=1}^{n} c_i \times \mathbb{1}[c_i \geq k]$), normalised by the maximum feasible number of disadvantages (i.e., $n \times d$). Alternatively, M may be interpreted as an average of the censored normalised MDSs, i.e., $(c_i \times \mathbb{1}[c_i \geq k])/d$. Like H, M is bounded between 0 and 1. When the disadvantage threshold is set at $k = d$, the following identity holds: $M(\mathscr{G}; d) = H(\mathscr{G}; d)$.

In order to understand the effectiveness of M in program evaluation, let us revisit our illustration comparing X and $\bar{X}_3$. For $k = 2$, the value of M for X or the pre-program mass is $5/12$ and the value of M for $\bar{X}_3$ or the post-program mass is $4/12$. The program, for $k = 2$, reduced the mass of disadvantages by $1/12$ in absolute terms or by 20% in relative terms. Clearly, M captures the program's positive impact that is missed in H.

Our primary outcome measure for multidimensional impact evaluation is M, but we also analyse 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 other hand, if the program primarily eliminates disadvantages among those with high MDSs but does 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 $X, \bar{X}_1$ and $\bar{X}_2$. Suppose, $k = 2$. The pre-program mass, incidence and intensity for X are $5/12$, $1/2$ and $5/6$, respectively. For $\bar{X}_1$, they are $1/6$, $1/4$ and $2/3$, respectively. Thus, a 60% reduction in M is accompanied by a 50% reduction in H and a 20% reduction in A. Let us now look at $\bar{X}_2$, where the mass, incidence and intensity are $1/4$, $1/4$ and 1, respectively. Unfortunately, in this case, the 40% reduction in M is accompanied by a 50% reduction in H, but a 20% increase in A.

2.2 Assessing distributional impact

So far, we have illustrated the framework using 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.[9] Let us revisit the illustration involving X, $\bar{X}_1$ and $\bar{X}_2$. First, consider $k = 1$. Three units in X experience one or more disadvantages and the pre-program mass is $1/2$. In $\bar{X}_1$, two units experience one or more disadvantages and the associated post-program mass is $1/4$. The post-program mass in $\bar{X}_2$ is also $1/4$. In both cases,

[9]The concept is analogous to poverty dominance (see, Atkinson, 1987; Foster and Shorrocks, 1988; Ravallion, 1994). For poverty dominance discussions in the counting framework, see Alkire et al. (2015, pp. 236–237).

the program has reduced the masses by 50%. Now, consider $k = 3$. The pre-program mass is $1/4$ for X, but the post-program masses for $\bar{X}_1$ and $\bar{X}_2$ are 0 and $1/4$, respectively. For $k = 3$, the program exhibits a positive impact when $\bar{X}_1$ is obtained from X, but does not show any change when $\bar{X}_2$ is obtained from X. Hence, if $\bar{X}_2$ is obtained from X, then the program cannot be considered inclusive because the unit that needs the most intervention did not benefit from the program's overall positive impact.

3 Philippines' Conditional Cash Transfer program

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 enrolment and attendance rates among children; (c) reduce the incidence of child labour; and (d) raise the average food consumption expenditure of poor households (DSWD, 2012).

Beneficiary households get two types of cash grants, released every two months: an education grant and a health grant. The education grant is ₱300 per month or ₱3,000 per year for each school-age child of 14 years or younger, for a maximum of three beneficiary children per household.[10] The education grant is expected to cover schooling expenses and to compensate families for possible income losses due to the schooling conditionality. The health grant is ₱500 per month or ₱6,000 per year. All identified beneficiaries are entitled to this grant, which aims to improve food consumption. The maximum overall grant that each household is entitled to receive is ₱15,000 per year.[11] The actual amount of grants that a beneficiary household receives depends on its composition and on its compliance with program conditionalities (Table A1). Teachers and local health workers monitor and verify compliance with these conditionalities.

The program follows a phased-in implementation design. Areas with the highest incidence of poverty based on 2006 poverty statistics are prioritised in 2008 but by 2010, 4Ps is initiated in all provinces. **Beneficiary households** are identified as follows. First, a household is identified as poor if its predicted income, estimated through a proxy means test (PMT), falls

[10]Program components described here are those applicable for the period covered in this evaluation study (2008 to 2011). For more details, see Fernandez and Olfindo (2011). At the time of data collection, the exchange rate was approximately US$1 = ₱45.

[11]The amount is around 15% of the poverty line income when 4Ps was initiated in treatment areas. Based on the 4Ps grants data, the average grant amount received by treatment households between January 2009 and November 2011 was ₱9,022 per year.

below the required poverty threshold.[12] 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 to validate their information and to formalise their program enlistment. In sum, eligible 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 as eligible at a village assembly.

The Philippine government considers 4Ps to be a major contributor to recent poverty reduction. The program is claimed to have led to an 82% increase in average income among the bottom three deciles of the population and to a decline in monetary poverty incidence from 26.3% to 21.6% between 2009 and 2015 (NEDA, 2017). Program evaluation reports show that the program improved outcomes and reduced non-compliance rates in different indicators. Onishi et al. (2013a), for instance, find positive impacts on school enrolment of children aged 3–11 years and on the nutritional status of children aged 6–36 months. Likewise, Onishi et al. (2013b) detect increases in the consumption of food and non-food items. Meanwhile, Orbeta et al. (2014) observe improvements in school enrolment among children aged 12 to 15 years, in deliveries in health facilities and in spending for education.

4 Effectiveness of 4Ps in alleviating joint non-compliances

The objective of 4Ps is to reduce poverty through compliance with program conditionalities. The first concern then is whether 4Ps is successful in reducing these non-compliances. Several studies examined the impact of 4Ps on non-compliance rates for different indicators separately, but none looked at the program's impact on joint non-compliances. Our aim in this section is to examine whether 4Ps reduced multiple non-compliances among indicators that are directly related to program conditionalities.

4.1 Data and experimental design

Admirably, 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

[12] Due to 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).

collects samples for both **randomised control trial** (RCT) and **regression discontinuity design** (RDD) evaluation. In this paper, we utilise only the first wave of the RCT evaluation survey. We prefer an RCT evaluation survey over an RDD evaluation survey because an RDD evaluation only allows us to capture localised treatment effects close to the program threshold. This may fail to capture the program's impact on the households in greatest need of intervention. Meanwhile, the second wave of the RCT evaluation survey is infeasible for our purpose because the control households from the first RCT survey are incorporated into the program from February 2012 onward.

The RCT survey follows a cluster randomised trial design, where treatment assignment is determined at the village level. In October 2008, eight municipalities were chosen to represent the poorest municipalities in the poorest provinces, and 130 clusters or villages were randomly drawn from these municipalities. Half of these villages were assigned to treatment. Program implementation in treatment areas commenced in January 2009, and the first wave of impact evaluation survey was carried out between October and November 2011. A total of 1,418 sample households were surveyed—704 from treatment and 714 from control villages. Treatment assignment was credibly implemented as no household from the control villages received 4Ps benefits according to the beneficiary database.[13]

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% of the sample households in the survey do not have any program-eligible member, potentially due to changes in household compositions between the time of the household assessment in 2008 and the time of the first wave survey in 2011. Our analysis is thus based on the 1,290 sample households, each with at least one program-eligible member. The distribution of program-eligible households in the treatment and control groups for our analysis is 641 and 649, respectively (Table A2).

4.2 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. Thus, we consider the eligible households residing in treatment villages as **treated** and the eligible households in control villages as **controls**,

[13]Meanwhile, 4% of 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 opted out of the program or were dropped from the list of eligible households during community validation as inclusion errors (Onishi et al., 2013a).

regardless of actual program status. In the literature, our approach is referred to as estimating **intent-to-treatment** (ITT) effect, or the average potential impact of offering the program. We thus capture the change in outcomes among the eligible households given the opportunity to participate in the program and not among the actual participants.[14]

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

$$y_i = \alpha + \tau p_i + \mathbf{x}_i \beta + \epsilon_i; \tag{4}$$

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, τ estimates the program's ITT effect, $\mathbf{x}_i$ is a vector of covariates related to household i that we control for, β is the vector containing coefficients of covariates in $\mathbf{x}_i$ and ϵ_i is the error term. A negative estimated value of τ reflects an improvement in the outcome variable and vice versa.

In this study, we consider a non-compliance as a disadvantage. We want to estimate the program's impact on the **masses of multiple non-compliances** for different non-compliance thresholds (k). Thus, our main outcome variable is the censored normalised multiple non-compliance score, i.e., $y_i = c_i/d$ if $c_i \geq k$ and $y_i = 0$ if $c_i < k$. We will estimate the program's impact on the incidence of non-compliances for different indicators as well as the incidence of multiple non-compliances for different k. Each outcome variable for evaluating the impact on incidences is a binary variable, such that $y_i = 1$ if household i experiences non-compliance (or experiences multiple non-compliances in the case of H) and $y_i = 0$ otherwise. We use linear probability models to estimate Equation 4 for the impact on incidences.[15]

We implement balance tests (a) by running a linear regression of each baseline covariate on treatment assignment, accounting for the cluster-randomised nature of the data and municipality fixed effects, and (b) by running a joint test of orthogonality of the baseline covariates against the treatment indicator. Our balance test results (Tables A3 and A4) does not show significant differences in the baseline covariates across treatment and control groups. The joint test also cannot reject the null hypothesis that the covariates are insignificant in predicting participation. Since a full baseline survey was not conducted, we are only able to test demographics and household characteristics used for computing the PMT incomes.[16] Still,

[14] For a critical review of anti-poverty program evaluation methods, see Ravallion (2007).
[15] To check the robustness of our findings, we compute marginal effects using probabilistic models. These alternative models produce similar analytical conclusions.
[16] To improve the efficiency of our estimates, we include baseline household characteristics that may affect

we have no reason to doubt that potential outcomes are 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 implementers.

4.3 Indicators and sample selection

We select five indicators that are directly targeted by the program. Table 1 presents the indicators and the non-compliance criteria, drawn closely from the program's conditionalities in Table A1. Our selection incorporates at least one indicator from each of the relevant target populations: school-age children (3–14 years old), children 0–5 years old, and women of reproductive age.[17]

Table 1: Indicators and non-compliance criteria for studying the 4Ps impact on multiple non-compliances

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

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 visits and birth delivery is available for female household members who are currently pregnant and have had a live birth in the past five

the variability of our outcomes. For instance, given that a larger number of program-eligible members may make a household more likely to experience non-compliance in a larger number of indicators, we control for the number of 4Ps-eligible members at baseline when estimating impact. We also include some village characteristics that are taken from the impact evaluation survey. The set of covariates for each outcome is detailed in each corresponding table of results.

[17] We could not include two other indicators—immunisation and post-natal care. Vaccination details are particularly challenging to recall and are taken from immunisation cards, which are presented for only around 24% of children aged 0–5. The post-natal care indicator suffers from similar significant missing data issues.

years. However, considering the program's exposure from January 2009 to September 2011, we only take into account births that are delivered from October 2009 onward. 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).

As mentioned, we focus our analysis on the 1,290 households with at least one program-eligible member. Yet, our multidimensional evaluation exercise requires us to look at compliance profiles of every household across all indicators simultaneously. In the present context, indicators each have respective applicable populations and so not all households have program-eligible member(s) in every indicator. For instance, only 25% of 4Ps-eligible sample households have at least one program-eligible member for the birth-delivery indicators, whereas 98% of 4Ps-eligible sample households have at least one member for the attendance indicator (Panel I of Table A2). Similarly, more than 90% of all 4Ps-eligible sample households have eligible members in at least two indicators, but less that 30% of all sample households have 4Ps-eligible members in four or more indicators (Panel II of Table A2).

Our multidimensional impact evaluation exercise thus entails a crucial trade-off. We may restrict our attention to sample households that have eligible members for all five selected indicators. Alternatively, we may consider all sample households with eligible member(s) in at least one indicator. The former option leads to a drastic reduction in sample size to merely 243 households (Table A2), which severely reduces the statistical power of our analysis and causes a loss of representativeness.

To elucidate the loss of representativeness, we divide the sample of all households with eligible member(s) in at least one indicator (Sample A) into a sample with eligible member(s) in at least one but less than five indicators (Sample B) and a sample with eligible members in all five indicators (Sample C). In Panel I of Table 2, we present the incidences of non-compliance for all five indicators in 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 non-compliance profiles in each sample, where the distribution for Sample C appears to be vastly different from the distribution of Sample B and thus from Sample A.

Considering the entire sample of 1,290 households certainly allows us to capture the impact of 4Ps without causing a loss of representativeness, but implicitly treats a household without any eligible member in an indicator to be compliant in that indicator.[18] It is infeasible for

[18]This approach is common for cross-country and inter-temporal comparisons in multidimensional poverty analysis (See, UNDP, 2010; Alkire and Santos, 2014; Alkire et al., 2017).

Table 2: Proportion of households with non-compliances 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***

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.
Notes: Under column headings Sample A, Sample B and Sample C, the left sub-column reports the number of sample households and the right sub-column reports the proportion of households. The final column (B — C) reports the difference of proportions between Sample B and Sample C and statistical significance.

every household, under this option, to be non-compliant in all five indicators, which may play a crucial role when targeting households by affecting inter-household comparability. Since we do not conduct any targeting exercise, however, such comparability is not a concern for our analysis. Thus, we primarily conduct our analysis on the entire sample of 1,290 households, but we verify the robustness of our findings for the sample of 243 households.[19]

4.4 4Ps impact on multiple non-compliances

The top half of Table 3 presents the estimated impacts on the incidences of non-compliance for the five indicators and the bottom half of the table presents the estimates on the masses of multiple non-compliances (M) for different non-compliance thresholds (k). We additionally report the estimated impacts on the incidences (H) and intensities (A) of multiple non-compliances. 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% confidence interval of the estimate, the counterfactual mean and the correspond-

[19]For balance tests for Sample A and Sample C across treatment and control groups, see Tables A3 and A4, respectively.

ing sample size, respectively. We control for household characteristics, municipality-level fixed effects and village-level variables (supply-side factors) that may affect the variability of program impact estimates.

Table 3: Estimates of 4Ps' impact on non-compliances 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)
	⟨1,266⟩	⟨752⟩	⟨1,104⟩	⟨394⟩	⟨322⟩
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)
	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩
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)	
	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	
A	−0.177***	−0.037	0.093	0.060	
	[−0.273, −0.080]	[−0.146, 0.071]	[−0.056, 0.243]	[−0.113, 0.233]	
	(2.106)	(2.708)	(3.388)	(4.246)	
	⟨1,060⟩	⟨653⟩	⟨313⟩	⟨106⟩	

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.
Notes: 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, the 90% confidence interval, the control mean and the number of observations are reported in the square brackets, in the parentheses and in the angular brackets, respectively. Baseline control variables include the household head's age and completed years of education and the number of program-eligible members in the household. 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. We also included municipality-level fixed effects. 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.

In the top half of Table 3, the sample size for each indicator corresponds to Panel I of Table A2. The program significantly improved three child-related indicators—attendance, health visit and deworming, by the magnitudes of 11.5, 11.8 and 7.3 percentage points, respectively. The health visit and deworming indicators are highly program-specific and so positive impacts show that the conditionalities are effective in inducing household behavioural changes. We are unable, however, to statistically detect changes in the incidences of non-compliances for the prenatal and birth delivery indicators.

In the bottom half of Table 3, we present the program's impact on the masses of multiple non-compliances. Overall, when we consider the households with one or more non-compliances ($k = 1$), we observe that the mass has fallen statistically significantly by 0.051 points or by 14.2% in relative terms. Decomposing the mass across the incidence and intensity of multiple compliances, we observe that the reduction in the mass is accompanied by a reduction in the incidence from 85.4% by 6.1 percentage points. At the same time, the intensity of multiple non-compliances of those experiencing one or more non-compliances is lower by 0.177 points on average, which is equivalent to slightly less than a fifth of an indicator.

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

Our conjecture is supported by the findings for $k = 4$ and $k = 5$. Even though the mass is lower by 13.3% for $k = 4$, this reduction is not statistically significant. The reduction in the corresponding incidence is also around 18% 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. Similar narrative unfolds for $k = 5$, where it is sufficient to interpret the change in the mass as $M = H$, but the sample size is too small for any meaningful conclusion. Still, 7.5% (i.e., 8.8% - 1.3%) or around 350,000 households experienced four or more non-compliances even after around two years of 4Ps exposure.

We thus observe a partial positive impact of 4Ps on multiple non-compliances. The program reduced non-compliances among households with three or fewer non-compliances by inducing desired behavioural changes through cash and conditionalities. The program, nevertheless, does not appear to have any impact on households experiencing a larger number of non-compliances.[20]

5 Is 4Ps inclusive?

Our findings in the previous section suggest that 4Ps left those with four to five non-compliances behind. Although our sample households are from the poorest areas and are identified as income poor, are the households with four to five non-compliances the poorest among them? If they are poorer than the rest of the households, then we may argue that 4Ps has not been inclusive, at least within the period under study. To examine this, we select indicators that capture various forms of deprivations but which are not directly targeted through 4Ps conditionalities. We then explore whether the households with four to five non-compliances are poorer than rest of the beneficiary households. Finally, we investigate whether the overall impact of the program has been shared by the households with four to five non-compliances.

5.1 Program's impact on deprivations

We select five indicators, chiefly based on three considerations. First, the selected indicators are related to program objectives but are not directly be targeted by 4Ps conditionalities. Second, each indicator can reflect changes in deprivations over the relatively short program exposure period 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 mostly 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.

A household is considered to be deprived in an indicator if the household fails to meet a subsistence standard or deprivation criterion for that indicator. The selected indicators and their deprivation criteria are listed in Table 4. Our first two indicators—consumption and hunger—are aligned with the program's objective of raising the average food consumption of poor households through health grants. For the hunger indicator, we avoid considering 'one occasion of hunger' as a reflection of potential deprivation because a single occurrence may be either due to recall or measurement bias or due to any other external shocks unrelated to

[20]In Table A5, we present the impact estimates based on the 243 sample households with eligible members in all five indicators. Most estimates are statistically insignificant due to the small sample size, but the absolute reduction in the mass for $k = 4$ is less than a 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 $k = 2$.

Table 4: 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 required subsistence level
Hunger	Household members had experienced hunger on more than one occasion in the past three months
Nutrition	Household has at least one child aged 0–5 years old, whose weight-for-age is two standard deviations lower than the median child growth standards
Dropout	Household has at least one child aged 3–14, who is not attending school
Savings	Household does not have a savings account or any other financial instrument

deprivation. The third indicator—nutrition—captures direct health deprivation within the household through child undernourishment as assessed by the World Health Organisation's child growth standards for weight-for-age.

The fourth indicator—dropout—may appear to be the same as the attendance indicator that we have used for our analysis in Section 4, but the dropout indicator is not directly targeted by program conditionalities. The program has a rather stricter criterion, which not only requires a child to be enrolled (the complement of dropout) but also requires at least an 85% 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 also the program does not prevent households from using the grant to enrol all of their children. The fifth indicator—savings—aims to reflect financial deprivation and identifies a household as deprived if the household does not have any savings account or any other savings instrument, such as a provident fund, life insurance or pre-need insurance. Maintaining a savings account aims to help beneficiaries smoothen their consumption or to open opportunities for other financial or enterprise assistance.

We examine the program's impact on the dincidence of each indicator separately and on their joint distribution or multidimensional poverty. To capture the impact on the joint distribution, we use the counting framework elaborated in Section 2. In this context, a deprivation is seen as a disadvantage. The mass of multiple deprivations is thus a reflection of multidimensional poverty, which is a product of the incidence and the intensity of multiple deprivations, divided by the number of indicators. We distinguish the notation used in this section from the notation used in Section 4 by assigning a prime. For instance, we denote the mass and the incidence of multiple deprivations by M' and H', respectively, and the deprivation threshold or the **poverty cut-off** by k'.

Table 5: 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					
	$k'=1$	$k'=2$	$k'=3$	$k'=4$	$k'=5$
M'	−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'	−0.118*	−0.120**	−0.036	−0.041	
	[−0.220, −0.016]	[−0.208, −0.031]	[−0.126, 0.054]	[−0.144, 0.062]	
	(2.153)	(2.670)	(3.383)	(4.132)	
	⟨1,194⟩	⟨820⟩	⟨364⟩	⟨116⟩	

$^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$. Source: Authors' own computations.
Notes: 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, the 90% confidence interval, the control mean and the number of observations are reported in the square brackets, in the parentheses, and in the angular brackets, respectively. Baseline control variables include the 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. We have also included municipality-level fixed effects. 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.

We use the regression specification in Equation 4 to estimate the causal impact of 4Ps, but the primary outcomes of interest here are **deprivation incidences** of the indicators in Table 4 and the masses of multiple deprivations for different poverty cut-offs. We use Sample A as specified in Table 2, and the set of controls includes selected household characteristics, municipality-level fixed effects and village-level characteristics, such as the number of schools.

In the top half of Table 5, we report impact estimates on deprivation incidences of the five indicators, where a negative estimate reflects a positive impact. The program's impact on deprivation incidences for four of the five indicators (consumption, hunger, underweight and savings) are small and statistically insignificant. The only statistically significant positive impact is observed for the dropout indicator. The underweight indicator, in fact, reflects a negative impact.

Does 4Ps exhibit a positive impact on multiple deprivations? In the lower half of Table 5, we present the program's estimated impact on the masses of multiple deprivations for five different poverty cut-offs: $k' = 1, \ldots, 5$. For $k' = 1$ and $k' = 2$, the statistically significant reductions in masses are 0.024 and 0.026 points, or 6% and 7.6%, respectively. In both cases, these reductions are driven by reductions in the corresponding intensities as the incidences do not change. A potential reason for observing no change in the incidences may be the prevalence of high deprivation for the savings indicator. For $k' = 3$, however, the statistically significant reduction in M' is 0.041 points (or 19.5%) and is accompanied by 5.8 percentage points reduction in the incidence. Even for $k' = 4$, both M' and H' declined statistically significantly by around 27%. In sum, 4Ps not only exhibits a positive impact on the overall masses of multiple deprivations (i.e., for $k = 1$), but it also exhibits a positive impact on the distribution of multiple deprivations.[21]

5.2 Is poverty reduction shared by the poorest?

To answer this question, we first explore whether the households with four to five non-compliances are associated with experiencing larger deprivations, on average, than the rest of the beneficiary households. For convenience, we denote the multiple non-compliance 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:

$$y_i^0 = \alpha_0 + \delta_0 T_i^{45} + \mathbf{x}_i^0 \beta_0 + \epsilon_i^0; \tag{5}$$

where, y_i^0 is the outcome of interest of household i, δ_0 estimates the difference in the averages of the outcome variable between those with four to five non-compliances and the rest of the households, β_0 is the vector containing coefficients of the covariates in $\mathbf{x}_i^0$ and ϵ_i^0 is the error term. Given that we are interested in the difference between the two groups prior to being treated, the estimates are based only on the control-group samples within Sample A.

We report the estimated differences in outcomes, $\hat{\delta}_0$, in Table 6. In the top half of the table, the $\hat{\delta}_0$ values reflect the differences in deprivation incidences for the five selected indicators.

[21] In Table A6, we present the impact estimates based on Sample C. Even though statistically insignificant due to the small sample size, the impact estimates are robust to our observations in Table 5.

Table 6: Comparison of deprivations between households with four to five non-compliances and rest of the beneficiary 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⟩

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.
Notes: Impact estimates are based on control group samples of 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 four to five non-compliances and the households with zero to three non-compliances. For each estimate, the 90% confidence interval, the mean outcome of the households with zero to three non-compliances and the number of observations are reported in square brackets, in parentheses, and in angular brackets, respectively. Baseline control variables include the 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. We have also included municipality-level fixed effects. Standard errors are robust to clustering at the village level. M' ranges between 0 and 1.

The households experiencing four to five non-compliances appear to be more deprived in all five indicators, but the differences are larger and statistically significant for the consumption and dropout indicators. Though statistically insignificant, the deprivation incidences in the hunger and underweight indicators are around 23–28% higher for households experiencing four to five non-compliances. Meanwhile, the bottom half of the table presents the differences in the masses of deprivations (M') for all five poverty cut-offs, $k' = 1, \ldots, 5$. We observe significantly lower masses for the households experiencing three or fewer non-compliances. For $k' = 1, \ldots, 4$, the estimated differences are between 0.190 and 0.240 points. Even for $k' = 5$, the estimated difference is 0.074 points. Thus, the households with four to five non-compliances are associated with experiencing larger joint deprivations or 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 by a small magnitude, of 4Ps. We answer this question by strictly focusing on the households that experience four to five non-compliances and examine the impact of 4Ps within this group. This exercise, however, is not straightforward because the selection of this group is not purely random and so the impact estimates may be subject to **selection bias**. To attenuate such bias, we resort to the Heckman selection procedure (Heckman, 1979).

Based on Equation 4, the multiple non-compliance scores may be estimated as

$$c_i^* = \gamma_1 + \tau_1 p_i + \mathbf{x}_i \gamma + \varepsilon_i; \tag{6}$$

where p_i is the binary program assignment such that $p_i = 1$ if household i resides in the treatment areas and $p_i = 0$ otherwise, $\mathbf{x}_i$ is a vector of other covariates, γ is the vector containing coefficients of covariates in $\mathbf{x}_i$ and ε_i is the error term. Note that $T_i^{45} = 1$ whenever $c_i^* = \gamma_1 + \tau_1 p_i + \mathbf{x}_i \gamma + \varepsilon_i \geq 4$. Therefore, the relevant sample selection equation is defined as

$$T_i^{45} = 1[\gamma_1 + \tau_1 p_i + \mathbf{x}_i \gamma + \varepsilon_i \geq 4] \text{ or } T_i^{45} = 1\left[\gamma_1' + \tau_1 p_i + \mathbf{x}_i \gamma + \varepsilon_i \geq 0\right]; \tag{7}$$

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 non-compliances is estimated by the following regression specification:

$$y_i^{45} = \alpha_1 + \delta_{45} p_i + \mathbf{x}_i^1 \beta_1 + u_i; \tag{8}$$

where y_i^{45} is the outcome variable for household i such that $T_i^{45} = 1$, δ_{45} is the coefficient for the program assignment variable p_i, $\mathbf{x}_i^1$ is a vector of covariates, β_1 is the vector containing the coefficients for the covariates in $\mathbf{x}_i^1$ and u_i is the error term. The error terms ε_i in Equation 7 and u_i in Equation 8 are assumed to follow a bivariate normal distribution with zero means, standard deviations σ_ε and σ_u, and correlation ρ. If $\hat{\rho} = 0$, then there is no sample selection problem (Wooldridge, 2010, p. 805) and the impact may be independently estimated by Equation 8. However, if $\hat{\rho} \neq 0$, then there is sample selection problem and the impact should be estimated jointly by Equations 7 and 8.

Since we use the program assignment variable p_i in the sample selection equation (7) and also in Equation 8, the program's impact on each outcome among those experiencing four to five non-compliances cannot simply be estimated by $\hat{\delta}_{45}$. Instead, it should be estimated by $\hat{\tau}_{45} = \hat{\delta}_{45} + h(\hat{\rho}, \hat{\sigma}_\epsilon, \hat{\sigma}_u, \hat{\gamma}_1', \hat{\tau}_1, \hat{\gamma})$, 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.

Table 7: Estimates of 4Ps' impact on deprivations among households experiencing four to five non-compliances, controlling for selection bias

	Consumption	Hunger	Underweight	Dropout†	Savings†
$\hat{\tau}_{45}$	−0.213***	0.075	0.006	0.026	−0.099
	[−0.343, −0.083]	[−0.062, 0.213]	[−0.155, 0.166]	[−0.147, 0.200]	[−0.208, 0.009]
	{0.645}	{0.032}	{0.265}	{0.355}	{0.016}
	(0.667)	(0.263)	(0.456)	(0.912)	(0.912)
	⟨1,290⟩	⟨1,290⟩	⟨716⟩	⟨1,290⟩	⟨1,290⟩

Impact on the joint distribution of deprivations (M')

	$k'=1$	$k'=2$	$k'=3^{\dagger}$	$k'=4$	$k'=5$
$\hat{\tau}_{45}$	−0.035	−0.029	−0.031	0.044	−0.038
	[−0.117, 0.047]	[−0.115, 0.056]	[−0.168, 0.106]	[−0.073, 0.161]	[−0.103, 0.027]
	{0.379}	{0.000}	{0.000}	{0.029}	{0.032}
	(0.596)	(0.582)	(0.912)	(0.298)	(0.088)
	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩	⟨1,290⟩

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.
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 four to five 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, the 90% confidence interval, the p-value for the Wald test for rejecting the null hypothesis of $\hat{\rho} = 0$ (a rejection confirms the existence of sample selection problem), the control mean and the number of observations are reported in square brackets, in curly brackets, in parentheses, and in angular brackets, respectively. Control variables for the selection equation (7) include the household head's age and completed years of education, the number of program-eligible members, municipality fixed effects 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. Full regression results are reported in Tables A7 and A8. Control variables for Equation 8 include the household head's completed years of education, the number of program-eligible members, municipality fixed effects and village-level variables—the numbers of grade and high schools. Standard errors are robust to clustering at the village level. Values of M' range between 0 and 1.
† For the dropout and savings outcomes, the MLE process had convergence problems, and we used the two-step Heckman estimation process. The p-value reported in the curly brackets corresponds to the t-test for rejecting the null hypothesis that the estimated coefficient for the inverse Mills ratio is zero.

7).

We present our findings in Table 7, where in the top half we report the estimated impact on the deprivation incidences of five indicators and in the bottom half we report the estimated impact on the masses of multiple deprivations. Again, a negative estimate indicates that the poorer households benefited from the program. In fact, if the estimates are larger in magnitude compared to the corresponding estimates in Table 5, then the impact would appear to be relatively favourable to the poorest of the poor.

Only consumption has improved statistically significantly among households experiencing four to five non-compliances. 4Ps has induced a 21.3 percentage points reduction in the deprivation incidence for consumption—considerably larger than the corresponding magnitude of the overall impact observed in Table 5. We have also observed improvement in the savings indicator, albeit statistically insignificantly. Although the program did not induce behavioural changes among the poorer households enough to hurdle all conditionalities, the 4Ps cash grants provided a sufficient income boost to hurdle the government's subsistence threshold for consumption expenditure. The result on the dropout indicator among the poorer households is unsatisfactory because the program has induced an overall reduction in the incidence of dropout deprivation of 8.8 percentage points (Table 5). No significant change is observed in the deprivation incidence for the rest of the indicators.

Finally, we examine the program's impact on the masses of multiple deprivations. We observe from Table 5 that the overall masses for $k' = 1$ to $k' = 4$ declined significantly by 0.024–0.041 points. Among the households with four to five non-compliances, we do not observe any statistically significant changes in the corresponding masses for all k'. 4Ps appears to have a limited contribution in reducing deprivations and poverty among the poorest.

6 Concluding remarks

The heightened emphasis on interconnected solutions to poverty raises important questions. An important one is: Do anti-poverty programs reach beneficiaries who simultaneously encounter multiple disadvantages? Conventional evaluation exercises study programs' impact on targeted outcomes separately, even when these programs explicitly target disadvantages on multiple outcomes simultaneously. Our key objective in this paper has been to examine whether an impact evaluation exercise, which incorporates the joint distribution of disadvantages, provides additional insights that may be missed by conventional evaluation exercises.

We first justify the use of a well-known counting framework (Atkinson, 2003; Alkire and Foster, 2011) for evaluating changes in the distribution of multiple disadvantages. We then apply the framework to investigate a successful anti-poverty program. We particularly looked at the impact of 4Ps on beneficiary households between 2009 and 2011. For our analysis, we use the first wave of randomised control trial survey embedded within 4Ps.

Like other CCT programs, 4Ps aims to arrest poverty by inducing behavioural changes through compliance with various conditionalities. In our first empirical exercise, we observe that 4Ps effectively induced behavioural changes among beneficiaries who do not comply with a smaller number of conditionalities, but it is not successful in producing the same changes among those who do not comply with a larger number of conditionalities simultaneously. In our second empirical exercise, we find that the beneficiaries who do not comply with a larger number of conditionalities are associated with experiencing higher deprivations, on average, than the rest. And while 4Ps successfully reduced consumption deprivation among these poorer beneficiaries, we do not find any conclusive evidence of reductions in poverty overall.

In the context of 4Ps, our findings suggest that for the poorest families, the cash grants may only be enough to marginally improve their consumption but not sufficient to alleviate other associated deprivations. This particular observation emphasises the need to specifically examine how poorer families manage program compliance as well as the requirements for complementary interventions to ensure that they are not left behind. Moreover, all families receive the same amount for their 4Ps health grant, but the grant may not be enough for larger families with pregnant women or young children. Larger, and also poorer, families may be reallocating part of their grants to food and other pressing needs. It may thus be beneficial to consider adjusting the health grants based on household composition.

Our analysis in this paper enriches the evidence on how anti-poverty programs perform in terms of reducing multiple disadvantages. This contributes to the need to draw new insights on the design and implementation of anti-poverty programs, given the increasing emphasis on interconnected solutions in the Sustainable Development Goals era, where we aim to reduce poverty in all its dimensions by the year 2030.

References

Alkire, S. and Foster, J. (2011). 'Counting and multidimensional poverty measurement', *Journal of Public Economics*, vol. 95(7), pp. 476–487.

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

Alkire, S., Roche, J.M. and Vaz, A. (2017). 'Changes over time in multidimensional poverty: Methodology and results for 34 countries', *World Development*, vol. 94, pp. 232–249.

Alkire, S. and Santos, M.E. (2010). 'Acute multidimensional poverty: A new index for developing countries', OPHI Working Paper 38, University of Oxford.

Alkire, S. and Santos, M.E. (2014). 'Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index', *World Development*, vol. 59, pp. 251–274.

Alkire, S. and Seth, S. (2015). 'Multidimensional poverty reduction in India between 1999 and 2006: Where and how?', *World Development*, vol. 72, pp. 93–108.

Atkinson, A. (2003). 'Multidimensional deprivation: Contrasting social welfare and counting approaches', *Journal of Economic Inequality*, vol. 1, pp. 51–65.

Atkinson, A.B. (1987). 'On the measurement of poverty', *Econometrica*, vol. 55(4), pp. 749–764.

Bag, S. and Seth, S. (2018). 'Does it matter how we assess standard of living? Evidence from Indian slums comparing monetary and multidimensional approaches', *Social Indicators Research*, vol. 140(2), pp. 715–754.

Baird, S., McIntosh, C. and Özler, B. (2011). 'Cash or condition? Evidence from a cash transfer experiment', *Quarterly Journal of Economics*, vol. 126(4), pp. 1709–1753.

Council of Economic Advisers (2018). 'Expanding work requirements in non-cash welfare programs', Executive Office of the President of the United States, Washington DC.

Das, J., Do, Q.T. and Özler, B. (2005). 'Reassessing conditional cash transfer programs', *World Bank Research Observer*, vol. 20(1), pp. 57–80.

de Janvry, A. and Sadoulet, E. (2006). 'Making conditional cash transfer programs more efficient: Designing for maximum effect of the conditionality', *World Bank Economic Review*, vol. 20(1), pp. 1–29.

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

Dutta, P., Murgai, R., Ravallion, M. and van de Walle, D. (2014). 'Does India's employment guarantee scheme guarantee employment?', *Economic and Political Weekly*, vol. 47(16), pp. 55–64.

Evans, D.K. and Popova, A. (2017). 'Cash transfers and temptation goods', *Economic Development and Cultural Change*, vol. 65(2), pp. 189–221.

Fernandez, L. (2012). 'Design and implementation features of the national household targeting system in the Philippines', Philippine Social Protection Note 5.

Fernandez, L. and Olfindo, R. (2011). 'Overview of the Philippines' conditional cash transfer program: The Pantawid Pamilyang Pilipino Program', Philippine Social Protection Note 2, pp. 1–12.

Filmer, D. and Schady, N. (2011). 'Does more cash in conditional cash transfer programs always lead to larger impacts on school attendance?', *Journal of Development Economics*, vol. 96(1), pp. 150–157.

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

Foster, J.E. and Shorrocks, A.F. (1988). 'Poverty orderings', *Econometrica*, vol. 56(1), pp. 173–7.

García, S. and Saavedra, J.E. (2017). 'Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: A meta-analysis', *Review of Educational Research*, vol. 87(5), pp. 921–965.

Glassman, A., Duran, D., Fleisher, L., Singer, D., Sturke, R., Angeles, G., Charles, J., Emrey, B., Gleason, J., Mwebsa, W., Saldana, K., Yarrow, K. and Koblinsky, M. (2013). 'Impact of conditional cash transfers on maternal and newborn health', *Journal of Health, Population, and Nutrition*, vol. 31, pp. S48–S66.

Heckman, J. (1979). 'Sample selection bias as a specification error', *Econometrica*, vol. 47, pp. 153–161.

Hoffmann, R. and Kassouf, A.L. (2005). 'Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure', *Applied Economics*, vol. 37(11), pp. 1303–1311.

Loschmann, C., R.Parsons, C. and Siegel, M. (2015). 'Does shelter assistance reduce poverty in Afghanistan?', *World Development*, vol. 74, pp. 305–322, doi:https://doi.org/10.1016/j.worlddev.2015.05.022.

National Economic and Development Authority (2017). *Philippine Development Plan 2017-2022*, National Economic and Development Authority.

OECD (2007). 'Combating poverty and social exclusion', in *OECD Economic Surveys: France 2007*, Paris: OECD Publishing, doi:https://doi.org/10.1787/eco_surveys-fra-2007-4-en.

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

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

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

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

Pasha, A. (2016). 'Impact of cash grants on multidimensional poverty in South Africa', Discussion Paper 208, Courant Research Centre, Georg-August-Universität Göttingen.

Ravallion, M. (1994). *Poverty Comparisons*, Switzerland: Harwood Academic Press.

Ravallion, M. (2007). 'Evaluating anti-poverty programs', *Handbook of Development Economics*, vol. 4, pp. 3787–3846.

UNDP (2010). *Human Development Report 2010: The Real Wealth of Nations: Pathways to Human Development*, New York: Palgrave Macmillan.

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

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel data*, Cambridge, MA and London: MIT press.

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

Appendix

Table A1: Program conditionalities for receiving grants

<table>
<tr><th>Grant type</th><th>Applicable member</th><th>Conditionalities</th></tr>
<tr><td>Education</td><td>Children 3–14 years old</td><td>Enroll in school and maintain a class attendance rate of at least 85% per month</td></tr>
<tr><td rowspan="3">Health</td><td>Children 0–14 years old</td><td>For children 0–5 years old:
• Complete all required vaccinations
• Monthly weight monitoring and nutrition counseling for children aged 0–23 months old; bi-monthly for 24–72 months old

For children 6–14 years old in elementary level:
• Take deworming pills twice a year</td></tr>
<tr><td>Pregnant women</td><td>Need to satisfy all the following criteria:
• One pre-natal consultation each trimester
• At least one blood pressure and weight monitoring measurement in each trimester
• At least one breastfeeding counseling session prior to delivery and within the first six weeks after childbirth
• At least one family planning counseling session prior to delivery and within the first six weeks after childbirth
• Delivery by a skilled health professional
• At least one post-natal care within the first six weeks after childbirth</td></tr>
<tr><td>Mother or other designated guardian</td><td>Attend monthly family development sessions (lectures on nutrition, sanitation, reproductive health, responsible parenthood, among others)</td></tr>
</table>

Source: Department of Social Welfare and Development (2012).

Table A2: Total number and distribution of 4Ps-eligible sample households across treatment and control groups

	Total sample		Treatment		Control	
4Ps-eligible sample households	1,290	(100.0)	641	(100.0)	649	(100.0)
Panel I: Samples with 4Ps-eligible members in each indicator reported in Table 1						
Attendance	1,266	(98.1)	631	(98.4)	635	(97.8)
Health visit	752	(58.3)	381	(59.4)	371	(57.3)
Deworming	1,104	(85.6)	551	(85.9)	553	(85.3)
Prenatal visit	394	(30.5)	199	(31.0)	195	(30.0)
Birth delivery	322	(25.0)	165	(25.8)	157	(24.3)
Panel II: Samples with 4Ps-eligible members for at least one to five indicators						
At least one indicator	1,290	(100.0)	641	(100.0)	649	(100.0)
At least two indicators	1,261	(97.8)	627	(98.7)	634	(97.6)
At least three indicators	698	(54.1)	350	(54.5)	348	(53.7)
At least four indicators	346	(26.8)	179	(28.0)	167	(25.6)
All five indicators	243	(18.8)	130	(20.3)	113	(17.5)

Notes: Values in parentheses within each column are percentages out of the overall 4Ps-eligible sample households reported in the first row.
Source: Authors' own computations.

Table A3: Balance tests of baseline household characteristics for households with eligible member(s) in at least one indicator†

	T	C	Δ
PMT per capita income	8,975	9,275	−215.470
Household composition			
Family size	5.886	5.875	0.020
Number of pregnant members	0.055	0.039	0.017
Number of children 0–5 years old	1.282	1.190	0.099
Number of children 6–14 years old	1.721	1.764	−0.040
Educational attainment of household head			
No grade completed	0.076	0.077	−0.001
Some elementary	0.431	0.427	−0.002
Elementary graduate	0.226	0.206	0.015
Some high school	0.105	0.133	−0.024
High school graduate	0.098	0.102	−0.000
Some college	0.041	0.035	0.007
College graduate and higher	0.023	0.020	0.005
Housing amenities			
Owns house and lot	0.310	0.356	−0.035
With strong roof materials	0.304	0.347	−0.030
With light and salvaged roof materials	0.696	0.653	0.030
With strong outer wall materials	0.264	0.267	0.007
With light salvaged outer wall materials	0.736	0.733	−0.007
With electricity	0.412	0.399	0.025
With Level 3 water system	0.190	0.191	−0.002
With water-sealed toilet	0.301	0.305	0.003
Household assets			
Owns a television	0.187	0.205	−0.010
Owns a car/jeep/motorcycle	0.033	0.026	0.009
Owns a cellphone/telephone	0.056	0.074	−0.015
Owns a video player	0.066	0.092	−0.023
Owns a stereo	0.090	0.100	−0.011
Number of observations	641	649	

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.
Notes: Figures are based on 1,290 sample households with at least one 4Ps-eligible member. Columns T and C report the means of each variable for the samples assigned to treatment and control, respectively. Column Δ reports the coefficient from the regression of each baseline characteristic on the treatment indicator, with municipality fixed effects and standard errors clustered at the village level. We also regressed the binary treatment assignment on all baseline household characteristics to test joint significance of the covariates in predicting treatment. The p-value from this joint F-test is 0.2616.
† The balance test is based on Sample A in Table 2.

Table A4: Balance tests of baseline household characteristics for households with eligible members in all indicators†

	T	C	Δ
PMT per capita income	8,576	8,978	−519.078
Household composition			
Family size	6.415	5.965	0.371
Number of pregnant members	0.069	0.071	−0.005
Number of children 0–5 years old	1.785	1.593	0.196
Number of children 6–14 years old	1.815	1.699	0.084
Educational attainment of household head			
No grade completed	0.085	0.080	0.010
Some elementary	0.408	0.442	−0.042
Elementary graduate	0.231	0.204	0.025
Some high school	0.123	0.106	0.019
High school graduate	0.069	0.106	−0.038
Some college	0.046	0.035	0.014
College graduate and higher	0.038	0.027	0.013
Housing amenities			
Owns house and lot	0.400	0.336	0.055
With strong roof materials	0.346	0.363	−0.047
With light and salvaged roof materials	0.654	0.637	0.047
With strong outer wall materials	0.338	0.248	0.059
With light salvaged outer wall materials	0.662	0.752	−0.059
With electricity	0.385	0.416	−0.030
With Level 3 water system	0.292	0.204	0.050
With water-sealed toilet	0.285	0.257	0.026
Household assets			
Owns a television	0.138	0.221	−0.082
Owns a car/jeep/motorcycle	0.031	0.027	0.008
Owns a cellphone/telephone	0.054	0.053	−0.003
Owns a video player	0.046	0.062	−0.016
Owns a stereo	0.062	0.133	−0.082**
Number of observations	130	113	

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Source: Authors' own computations.

Notes: Figures are based on 243 sample households with 4Ps-eligible members in all indicators. Columns T and C report the means of each variable for the samples assigned to treatment and control, respectively. Column Δ reports the coefficient from the regression of each baseline characteristic on the treatment indicator, with municipality fixed effects and standard errors clustered at the village level. We also regressed the binary treatment assignment on all baseline household characteristics to test joint significance of the covariates in predicting treatment. The p-value from this joint F-test is 0.1842.

† The balance test is based on Sample C in Table 2.

Table A5: Estimates of 4Ps' impact on non-compliances for households with eligible members in all indicators

	Attendance	Health visit	Deworming	Prenatal	Birth delivery
	−0.099	0.019	−0.076	−0.057	−0.022
	[−0.212, 0.013]	[−0.059, 0.098]	[−0.169, 0.017]	[−0.161, 0.046]	[−0.121, 0.077]
	(0.522)	(0.796)	(0.699)	(0.381)	(0.690)
	⟨243⟩	⟨243⟩	⟨243⟩	⟨243⟩	⟨243⟩

Estimates of impact on the joint distribution of non-compliances

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
M	−0.047*	−0.048	−0.094**	−0.026	−0.006
	[−0.094, −0.000]	[−0.101, 0.005]	[−0.165, −0.023]	[−0.114, 0.062]	[−0.077, 0.065]
	(0.618)	(0.602)	(0.520)	(0.308)	(0.124)
	⟨243⟩	⟨243⟩	⟨243⟩	⟨243⟩	⟨243⟩
H	−0.026	−0.029	−0.144**	−0.031	
	[−0.055, 0.003]	[−0.087, 0.029]	[−0.236, −0.051]	[−0.130, 0.067]	
	(0.991)	(0.912)	(0.708)	(0.354)	
	⟨243⟩	⟨243⟩	⟨243⟩	⟨243⟩	
A	−0.186	−0.206	0.027	0.012	
	[−0.414, 0.043]	[−0.419, 0.007]	[−0.202, 0.256]	[−0.176, 0.200]	
	(3.116)	(3.301)	(3.675)	(4.350)	
	⟨238⟩	⟨218⟩	⟨157⟩	⟨85⟩	

$^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$. Source: Authors' own computations.
Notes: Results are based only on Sample C, as described in Table 2. 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, the 90% confidence interval, the control mean and the number of observations are reported in the square brackets, in the parentheses and in the angular brackets, respectively. Baseline control variables include the household head's age and completed years of education and the number of program-eligible members in the household. 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. We also included municipality-level fixed effects. 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.

Table A6: Estimates of 4Ps' impact on deprivations and poverty for households with eligible members in all indicators

	Consumption	Hunger	Underweight	Dropout	Savings
	−0.039 [−0.142, 0.063] (0.619) ⟨243⟩	−0.022 [−0.110, 0.065] (0.230) ⟨243⟩	0.036 [−0.077, 0.148] (0.402) ⟨242⟩	−0.043 [−0.152, 0.065] (0.345) ⟨243⟩	−0.022 [−0.096, 0.051] (0.867) ⟨243⟩

Estimates of impact on the joint distribution of deprivations

	$k'=1$	$k'=2$	$k'=3$	$k'=4$	$k'=5$
M'	−0.018 [−0.072, 0.036] (0.492) ⟨243⟩	−0.012 [−0.079, 0.055] (0.453) ⟨243⟩	−0.046 [−0.126, 0.034] (0.340) ⟨243⟩	−0.052 [−0.124, 0.019] (0.207) ⟨243⟩	−0.015 [−0.048, 0.018] (0.044) ⟨243⟩
H'	0.005 [−0.045, 0.054] (0.947) ⟨243⟩	0.034 [−0.061, 0.128] (0.752) ⟨243⟩	−0.051 [−0.162, 0.059] (0.469) ⟨243⟩	−0.062 [−0.147, 0.024] (0.248) ⟨243⟩	
A'	−0.113 [−0.377, 0.151] (2.598) ⟨232⟩	−0.199 [−0.400, 0.003] (3.012) ⟨191⟩	−0.126 [−0.319, 0.066] (3.623) ⟨110⟩	−0.074 [−0.243, 0.096] (4.179) ⟨53⟩	

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.

Notes: Results are based only on Sample C, as described in Table 2. 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, the 90% confidence interval, the control mean and the number of observations are reported in the square brackets, in the parentheses, and in the angular brackets, respectively. Baseline control variables include the 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. We have also included municipality-level fixed effects. 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.

Table A7: The Heckman regression estimates for households with four to five non-compliances

	Consumption	Hunger	Underweight	Dropout†	Savings†
Selection equation results					
Program assignment ($\hat{\tau}_1$)	−0.145	−0.141	−0.179	−0.143	−0.143
No. of 4Ps-eligible members at baseline	0.185***	0.187***	0.162***	0.186***	0.186***
Age of household head	−0.033***	−0.033***	−0.019***	−0.033***	−0.033***
Educational attainment of household head	−0.053***	−0.053***	−0.063***	−0.053***	−0.053***
No. of grade schools in the village	0.038	0.038	0.033	0.039	0.039
No. of high schools in the village	0.019	0.025	0.020	0.019	0.019
No. of health centers in the village	0.092	0.117	0.012	0.096	0.096
No. of doctors in the village	0.017	0.006	0.029	0.017	0.017
No. of midwives in the village	−0.018	−0.008	−0.003	−0.008	−0.008
Outcome equation results					
Program assignment ($\hat{\delta}_{45}$)	−0.199**	0.104	0.048	0.050	−0.064
No. of 4Ps-eligible members at baseline	0.045	−0.028	0.020	−0.071**	−0.035*
Educational attainment of household head	−0.031*	−0.021	0.007	−0.020	−0.012
No. of grade schools in the village	0.073*	−0.003	0.048	0.006	0.055*
No. of high schools in the village	−0.193***	−0.074	0.054	0.017	−0.057
Observations	1,290	1,290	716	1,290	1,290

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.
Notes: All estimated regressions in the top half correspond to Equation 7 and those in the bottom half correspond to Equation 8. The outcome indicators are in Panel I of Table 7. We control for municipality fixed effects, which are not reported in the table. Standard errors (not reported) are robust to clustering at the village level. Here we show that the variable educational attainment of the household head, and to a weak extent the variable number of 4Ps-eligible members at baseline, satisfy the exclusion restriction criteria for the Heckman selection correction procedure.
† For the dropout and savings outcomes, the MLE process had convergence problems and we used the two-step Heckman estimation process.

Table A8: The Heckman regression estimates for households with four to five non-compliances

	$M_0(k'=1)$	$M_0(k'=2)$	$M_0(k'=3)^{\dagger}$	$M_0(k'=4)$	$M_0(k'=5)$
Selection equation results					
Program assignment ($\hat{\tau}_1$)	−0.145	−0.138	−0.143	−0.141	−0.139
No. of 4Ps-eligible members at baseline	0.184***	0.173***	0.186***	0.186***	0.187***
Age of household head	−0.034***	−0.027***	−0.033***	−0.034***	−0.034***
Educational attainment of household head	−0.051***	−0.046***	−0.053***	−0.053***	−0.053***
No. of grade schools in the village	0.039	0.031	0.039	0.042	0.038
No. of high schools in the village	0.026	0.026	0.019	0.016	0.022
No. of health centers in the village	0.006	−0.047	0.096	0.082	0.109
No. of doctors in the village	0.010	0.001	0.017	0.016	0.011
No. of midwives in the village	−0.015	−0.015	−0.008	−0.002	0.001
Outcome equation results					
Program assignment ($\hat{\delta}_{45}$)	−0.004	0.031	0.006	0.072	−0.025
No. of 4Ps-eligible members at baseline	−0.019	−0.056***	−0.025	−0.014	0.000
Educational attainment of household head	−0.014	−0.003	−0.019	−0.013	−0.009
No. of grade schools in the village	0.030	0.006	0.063	0.017	−0.032*
No. of high schools in the village	−0.055	−0.079	−0.060	−0.105	0.007
Observations	1,290	1,290	1,290	1,290	1,290

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Source: Authors' own computations.
Notes: All estimated regressions in the top half correspond to Equation 7 and those in the bottom half correspond to Equation 8. The outcome indicators are in Panel II of Table 7. We control for municipality fixed effects, which are not reported in the table. Standard errors (not reported) are robust to clustering at the village level. Values of M' range between 0 and 1. Here we show that the variable educational attainment of the household head, and to a weak extent the variable number of 4Ps-eligible members at baseline, satisfy the exclusion restriction criteria for the Heckman selection correction procedure.
† For $M_0(k'=3)$ outcome, the MLE process had convergence problems and we used the two-step Heckman estimation process.