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

OPHI RESEARCH IN PROGRESS SERIES 55a

Evaluation of Programs with Multiple Objectives: Multidimensional Methods and Empirical Application to *Progresa* in Mexico

Ana Vaz*, Bilal Malaeb** and Natalie Naïri Quinn***

August 2019

Abstract
Development programs and policy interventions frequently have multiple simultaneous objectives. Existing quantitative evaluation approaches fail to fully accommodate this multiplicity of objectives. In this paper we adapt the multidimensional poverty measurement approach developed by Alkire and Foster (2011) to the estimation of treatment effects for programs with multiple objectives. We use the potential outcomes framework to show that differences in Alkire-Foster indices between treated and control samples correspond to average treatment effects estimates of outcomes of interest under experimental conditions, and develop further methods of analysis to explore these multidimensional treatment effects. We discuss issues of index design encountered in practice and provide an illustrative example. We apply the methods developed to evaluate the conditional cash transfer program *Progresa* in Mexico, finding significant multidimensional effects of the program. Further analysis shows that these treatment effects are driven mainly by impacts on school attendance and health visits, objectives that correspond directly to the conditions of the program. There is no evidence for heterogeneity of the treatment effects by the extent to which the beneficiary failed to achieve the objectives at baseline. This study complements the extensive literature on the evaluation of *Progresa* and other development programs, comprising studies that focus on particular objectives or outcomes of the program. We hope that the methods developed here will find wide application to the evaluation of programs with multiple objectives.

* Senior Research Officer at Oxford Poverty and Human Development Initiative (OPHI), University of Oxford. ana.vaz@qeh.ox.ac.uk.

** Postdoctoral Research Officer, Institute of Global Affairs, London School of Economics and Political Science. B.Malaeb@lse.ac.uk.

*** Senior Research Officer at Oxford Poverty and Human Development Initiative (OPHI), University of Oxford. natalie.quinn@qeh.ox.ac.uk.

This paper is part of the Oxford Poverty and Human Development Initiative's Research in Progress (RP) series. These are preliminary documents posted online to stimulate discussion and critical comment. The series number and letter identify each version (i.e. paper RP1a after revision will be posted as RP1b) for citation.

For more information, see www.ophi.org.uk.

Keywords: poverty; impact evaluation; interventions with multiple outcomes; conditional cash transfer.

JEL classification: I32, O12, O22.

Acknowledgements:
We would like to thank participants in the OPHI workshop Multidimensional Impact Evaluation, the Sustainability and Development Conference, and OPHI seminars for helpful comments. We are also grateful to the Economic and Social Research Council for financial support.

Funding information: This paper is supported by Economic and Social Research Council, grant number ES/N01457X/1.

Citation: Vaz, A., Malaeb, B. and Quinn, N.N. (2019). 'Evaluation of programs with multiple objectives: Multidimensional methods and empirical application to *Progresa* in Mexico', OPHI Research in Progress 55a, University of Oxford.

1. Introduction

The 2030 Agenda for Sustainable Development reflected two shifts that have been taking place in the context of social policy. First, it recognized that poverty has a multidimensional nature (Sustainable Development Goal 1); second, it emphasized intersectoral linkages and the need to adopt integrated approaches to policymaking (UN General Assembly, 2015). In this context, it is expected that more and more social protection schemes and poverty reduction programs (generally, 'policy interventions') will adopt multidimensional approaches. They will aim to simultaneously tackle different deprivations, drawing on the synergies between the different areas of interventions and enhancing the effectiveness of the interventions (Skoufias, 2005). To date, one of the most notable examples of this type of policy is the conditional cash transfer programs that aim to improve investment in children's health and education. Another example is the Millennium Villages that aim to promote an integrated approach to rural development by improving access to safe water, sanitation and other important infrastructure, and focusing on environmental sustainability.

Also in recent years, there has been a growing emphasis placed on the evaluation of policy interventions as part of a global trend towards *evidence-based policy making* (Gertler et al., 2016). One of its classic examples is the evaluation of the pioneer conditional cash transfer program Progresa in Mexico, one of the first national poverty alleviation programs to adopt a design that would allow a rigorous assessment of its impact. Rigorous evaluation is important to provide evidence of what works and it is 'an accountability tool at the end of a project cycle' (Kremer, 2010 at UN Summit in New York).

In this context, it seems natural that the evaluation of policy interventions with a multidimensional approach should also be multidimensional. In this paper we adapt and apply the approach to multidimensional poverty measurement developed by Alkire and Foster (2011), henceforth AF, to the evaluation of policy interventions with multiple objectives, illustrating its value in this context.

Consider the assessment of social programs and policies that include interventions in different domains (e.g. education, health, etc.) and/or are expected to have multiple simultaneous effects (e.g. increase school attendance, reduce child labor, reduce child malnutrition, etc.). Standard impact evaluation practice is to measure the average effect of the program on each of the outcome indicators separately (e.g. rate of school enrolment, rate of child labor, children's average calorie intake, etc.), making appropriate adjustments to avoid multiple inference errors. This approach, however, does not allow us to conclude that the program's goals were achieved simultaneously; it does not enable us to see if the children who started attending school are the same as those who stopped working, or the same who started eating more. Furthermore, in the case of untransformed outcome indicators, a positive average effect does not give us any information on whether the individuals that experienced an improvement were the ones who were most deprived or not. For instance, it is possible that an increase in calorie intake could have benefited especially those who already had a sufficient level of calorie intake. We argue that the AF approach provides a more direct measure of the program's overall performance. By focusing the evaluation on the individual instead of on the indicator, it allows us to have a better perception of whether the program is lifting people out of poverty, instead of having impacts on scattered indicators. This methodology allows us to monitor the impact of the program on the joint distribution of deprivations. By summarizing the impact in different domains into one outcome index, we avoid the multiple comparisons problem and the loss of power associated with standard correction techniques. Finally, the AF methodology has an advantage in terms of communication as it allows us to summarize the impact in different domains into one single number, instead of a dashboard of indicators.

A different approach that also utilizes a summary index, increasingly applied in the evaluation of policy interventions with multiple outcomes, is to combine multiple indicators into a single *linear* index following Anderson (2008). Anderson's index follows O'Brien (1984) and yields the best linear unbiased estimator if each of the indicators were interpreted as a noisy indicator of some underlying latent variable. While there are compelling reasons to adopt this approach if the researcher's objective is to maximize statistical power, the index so created may bear little relation to the objectives of the policy intervention. A positive result would demonstrate that the intervention had *some* effect, but not necessarily the *intended* effect. We argue that the AF approach, with indicators and weights chosen to reflect the objectives of the policy intervention, is in many contexts a more appropriate approach to summary index construction for impact evaluation.

Some poverty reduction programs already base their targeting (at least partially) on multidimensional indices of unmet basic needs / quality of life (Solidarity Chile, Families in Action in Colombia, Tekopora in Paraguay and the definition of Below Poverty Line in India). However, to our knowledge, none of these indices is built based on the AF methodology.[1] In the

[1] Most of these indices, if not all, are vulnerable to two main methodological critiques: cardinality and substitutability. See Alkire and Seth (2008) for more details on these critiques in the scope of the 2002 Below Poverty Line methodology.

literature, however, there has been an increasing interest in linking the AF methodology with the targeting of poverty reduction interventions (Bouillon and Yanez-Pagans, 2011; Robano and Smith, 2012; Alkire and Seth, 2013; Azevedo and Robles, 2013).

Despite these developments in targeting, the potential of AF methodology to evaluate the performance of poverty reduction programs remains largely unexplored. Robano and Smith (2013) have briefly illustrated how multidimensional measurement technology can be used for impact assessment using data from a poverty reduction program in Bangladesh, while Loschmann, Parsons and Siegel (2014) have used the AF methodology to assess the overall impact of shelter assistance in Afghanistan.

This paper makes two main contributions. First, it illustrates the advantages of using the AF multidimensional measurement methodology to assess the impact of policy interventions with multiple outcomes, in comparison to standard impact evaluation with a set of separate indicators or a statistical summary index. Second, this paper provides a detailed multidimensional framework for impact evaluation, including a complete battery of complementary analyses that can enhance understanding of the program's effects.

The paper is organized as follows. In section 2 we show how the Alkire-Foster approach may be adapted to evaluate policy interventions and programs with multiple objectives, and suggest further analyses that may be conducted to further explore the treatment effects found. In section 3 we discuss practical issues of measure design that must be considered when applying the methods proposed to real-world evaluation tasks. In section 4 we present a simple illustrative example to demonstrate the methods and explore their advantage over separate evaluation of a dashboard of indicators. In section 5 we apply the basic methods and further analyses to assess the impact of *Progresa* in Mexico. Section 6 concludes.

2. Methodological Framework

Alkire and Foster (2011; see also Alkire et al 2015) proposed an approach to multidimensional measurement of poverty, the AF method, that combines counting of deprivations in different dimensions to identify the poor[2] with an adaptation of the Foster, Greer and Thorbecke (1984) family of poverty measures to aggregate information across dimensions and across the poor. This approach has found many applications,[3] notably the Global Multidimensional Poverty Index (MPI) published in the Human Development Report since 2010 (Alkire and Santos, 2014), and the national multidimensional poverty indices that have been developed by many countries including Mexico (CONEVAL, 2010) and Colombia (Salazar et al., 2011). Another application of the AF method is the Women's Empowerment in Agriculture Index[4] (Alkire et al., 2013).

In this section we show that Alkire and Foster's (2011) approach (the AF method) may be straightforwardly adapted to the evaluation of policy interventions with multiple objectives. We formally show that under experimental conditions, treatment effects of interest may be estimated by the difference in AF indices between treated and control groups. We then describe several further analyses that may be implemented to explore the multidimensional character of the program impact.

2.1 Adaptation of the AF method to program evaluation

The AF method may be naturally adapted to program evaluation in the following context. Suppose that a program or policy intervention has $D > 1$ objectives that can be measured with D indicators of achievement x_d, $d = 1, \dots, D$.[5] These indicators need not have cardinal interpretation. Suppose that each objective d can be defined in terms of a minimum achievement threshold z_d for each target unit (individual, household, community, etc.) n, so the objective is achieved if $x_{nd} \geq z_d$.[6] Let w_d be the weight placed on objective d, such that $w_d \in \mathbb{R}^+$ and

2 Identification is based on a dual cutoff: a deprivation cutoff for each dimension to determine who is deprived in that dimension, and a poverty cutoff – the proportion of dimensions in which an individual needs to be deprived in order to be considered multidimensionally poor. This identification approach generalises the traditional union and intersection approaches.

3 The AF method is very flexible and can be employed in a variety of contexts and purposes by adopting different dimensions, indicators, deprivation cutoffs, weights and poverty cutoffs.

4 Developed jointly by the United States Agency for International Development (USAID), the International Food Policy Research Institute (IFPRI), and Oxford Poverty and Human Development Initiative (OPHI).

5 It may be that some objectives have more than one indicator of achievement. If multiple indicators are used for one objective, thresholds and weights must be chosen for each of the indicators. For clarity of exposition, we assume one indicator per objective here.

6 For example, a program that aims at ensuring minimum levels of nutrition and education by guaranteeing that all adults have a body mass index above 18.5 and at least 10 years of schooling.

$\sum_{d=1}^{D} w_d = 1$, and suppose that the overall goal of the program is to reduce the weighted sum of each target unit's missed objectives below a certain cutoff $k \in [0,1]$. (Or equivalently, to increase the weighted sum of attained objectives above $1-k$.)

We will refer to the program target unit as an individual, but the exposition would be analogous for any other target unit, for example household or community. Individual n will be evaluated as not attaining objective d if $x_{nd} < z_d$. Following the AF conceptual terminology, individual n may be referred to as *deprived* in indicator d; the proportion of people that are deprived in indicator d is the *uncensored headcount ratio* of indicator d.

A first possible outcome index at the individual level is the *weighted sum of missed objectives*, analogous to the *deprivation score* in multidimensional poverty measurement,

$$c_n = \sum_{d=1}^{D} w_d \, I(x_{nd} < z_d). \tag{1}$$

If $c_n \geq k$, individual n has not achieved the program's overall goal. A second possible outcome index is *individual status*, a binary indicator of whether the individual has failed to achieve the program's overall goal,

$$h_n = I(c_n \geq k). \tag{2}$$

This is analogous to the indicator of being multidimensionally poor, *poverty status,* in the multidimensional poverty measurement context.

While the overall goal for each individual is to reduce c_n below the cutoff k, the program evaluator is likely to want to take into account progress made toward that goal, for those who do not achieve it. This may be captured by a third possible outcome index at the individual level, the censored weighted sum of missed objectives or *censored deprivation score*,

$$c(k)_n = h_n c_n. \tag{3}$$

This index is equal to the weighted sum of missed objectives if the individual has not achieved the program's overall goal, while it is equal to zero if the individual has achieved the overall goal. Thus, it censors the missed objectives of those who have achieved the program's overall goal. The notation $c(k)_n$ distinguishes this censored index from the original sum of weighted missed objectives, c_n.

The program evaluator wishes to determine the effect of the program on one or more of these outcome indices, c_n, h_n and $c(k)_n$. Let $T_n = 1$ if individual n is exposed to the program, and $T_n = 0$ if they are not. Applying the potential outcomes framework (Rubin, 2005), let $c_n(1)$

represent the potential deprivation score for individual n if they are exposed to the program, and $c_n(0)$ represent the potential deprivation score for that same individual if they are not; potential individual statuses $h_n(1)$ and $h_n(0)$ and potential censored deprivation scores $c(k)_n(1)$ and $c(k)_n(0)$ are defined similarly.

The causal effects of the program for individual n are then $c_n(1)-c_n(0)$, $h_n(1)-h_n(0)$ and $c(k)_n(1)-c(k)_n(0)$. Of course, none of these is observable, as each individual is either exposed to the program or not, and the counterfactual is not observed. But it may be possible to estimate the average treatment effects,

$$ATE_c = \mathbb{E}\big(c_n(1)\big) - \mathbb{E}\big(c_n(0)\big), \tag{4}$$

$$ATE_h = \mathbb{E}\big(h_n(1)\big) - \mathbb{E}\big(h_n(0)\big) \text{ and} \tag{5}$$

$$ATE_{c(k)} = \mathbb{E}\big(c(k)_n(1)\big) - \mathbb{E}\big(c(k)_n(0)\big). \tag{6}$$

Suppose that exposure to the program is randomly assigned, so that $T_n \perp Y_n(1)$ and $T_n \perp Y_n(0)$ where $Y_n = c_n, h_n$ or $c(k)_n$, then $\mathbb{E}(Y_n(1)|T_n = 1) = \mathbb{E}\big(Y_n(1)\big)$ and $\mathbb{E}(Y_n(0)|T_n = 0) = \mathbb{E}\big(Y_n(0)\big)$. If outcomes are observed for a random sample $n = 1, \dots, N^T$ of individuals exposed to the program ('treated') and the Stable Unit Treatment Value Assumption holds then $\bar{Y}^T = \frac{1}{N^T}\sum_{n=1}^{N^T} Y_n$ is an unbiased estimator of $\mathbb{E}(Y_n(1)|T_n = 1)$ and if outcomes are observed for a random sample $i = 1, \dots, N^C$ of individuals not exposed to the program ('control') then $\bar{Y}^C = \frac{1}{N^C}\sum_{i=1}^{N^C} Y_n$ is an unbiased estimator of $\mathbb{E}(Y_n(0)|T_n = 0)$. We then have unbiased estimators for the average treatment effects (4), (5) and (6),

$$A\hat{T}E_c = \frac{1}{N^T}\sum_{n=1}^{N^T} c_n - \frac{1}{N^C}\sum_{i=1}^{N^C} c_n, \tag{7}$$

$$A\hat{T}E_h = \frac{1}{N^T}\sum_{n=1}^{N^T} h_n - \frac{1}{N^C}\sum_{i=1}^{N^C} h_n \text{ and} \tag{8}$$

$$A\hat{T}E_{c(k)} = \frac{1}{N^T}\sum_{n=1}^{N^T} c(k)_n - \frac{1}{N^C}\sum_{i=1}^{N^C} c(k)_n. \tag{9}$$

Returning to the framework provided by Alkire and Foster (2011), we may observe that (8) is directly analogous to the difference in headcount ratio or *incidence* of poverty between the treated and control samples, $A\hat{T}E_h = H^T - H^C$. Similarly, (9) is analogous to the difference in adjusted headcount ratio or *multidimensional poverty index* M_0 between the treated and control samples, $A\hat{T}E_{c(k)} = M_0^T - M_0^C$.

The adjusted headcount ratio, M_0, corresponds to the ratio of the weighted sum of missed objectives for those in the sample who do not attain the program's overall goal to the potential maximum if none of the individuals attained any of the objectives.[7] As this index most fully captures the extent to which the program's multiple objectives are achieved, we propose $A\hat{T}E_{c(k)} = M_0^T - M_0^C$ as the primary treatment effect for evaluating the impact of a program with multiple objectives.

In practice, of course, estimation of the treatment effects may not be as simple as computation of observed differences between treated and control samples. Where treatment is randomly assigned, statistical power may be improved by use of baseline data or repeated observations (McKenzie 2012) and complexities such as imperfect compliance and spillovers must be taken into account. Where assignment is not random, selection bias is an issue and an appropriate quasi-experimental approach must be found. The methods most commonly applied to evaluate development programs are surveyed in Ravallion (2007).

We conclude by noting that Alkire and Foster (2011) propose a wider class of multidimensional poverty indices,

$$M_\alpha = \frac{1}{N}\sum_{n=1}^{\mathrm{N}} I(c_n \geq k) \sum_{d=1}^{D} w_d \left(1 - \frac{x_{nd}}{z_d}\right)^\alpha I(x_{nd} < z_d)\ , \qquad (10)$$

where the parameter $\alpha \geq 0$. Application of indices with $\alpha > 0$ relies on all indicators having a cardinal interpretation; as this rarely holds in practice, we restrict attention to $\alpha = 0$ as developed above.

2.2 Further analyses of multidimensional impact

Recall that under random assignment to treatment and SUTVA our main treatment effect of interest, $ATE_{c(k)}$ (6) is estimated by the difference in Alkire and Foster's (2011) adjusted headcount ratio M_0 between treatment and control groups. Within the AF framework, there are a number of analyses that can help us to better understand the multidimensional nature of poverty. Here we show how they can be adapted to explore a multidimensional treatment effect. For simplicity of exposition we continue to assume that assignment to treatment is random and SUTVA applies, so that observed treatment-control differences are unbiased estimators of the

7 If all targeted individuals achieve the program's overall goal this measure will take the value zero for the targeted group.

treatment effects; in practice, evaluation methods that are appropriate to the context would need to be applied to estimate the objects of interest.

a) Decomposition of treatment effect by incidence and intensity

Alkire and Foster (2011) show that the adjusted headcount ratio M_0 may be expressed as the product of the headcount ratio H and a measure of the *intensity* of poverty,

$$A = \frac{1}{N_p} \sum_{n=1}^{N} c(k)_n \quad (11)$$

where $N_p = \sum_{n=1}^{N} h_n$ is the number of the poor. In the present context, this corresponds to the average weighted sum of missed objectives for those who fail to achieve the program's overall goal, which we may describe as *intensity of missed objectives*. While intensity may be evaluated for the treatment and control groups, its difference between these groups cannot be interpreted as an estimate of an average treatment effect, because the denominator varies according to the number of individuals who fail to achieve the overall goal.

However, suppose that a policy intervention with multiple objectives achieves a positive effect, so $M_0^T < M_0^C$. The difference in M_0 between the groups may arise from two different effects: a increase in the number of individuals who achieve the overall goal of the program (a reduction in H) and a reduction in the intensity of missed objectives among those who nevertheless fail to achieve the overall goal (a reduction in A). Ideally we would be able to decompose the overall reduction in M_0 into these two separate effects.

Unfortunately, in practice we can never observe the counterfactual outcome for a specific individual; we can never know exactly what the outcome would have been for an individual who was treated, had they not been treated. So it is impossible to carry out an exact decomposition. This is exactly analogous to the decomposition of changes in poverty over time explored by Alkire, Roche and Vaz (2017) in the case where panel data is not available, but repeated cross-sections are. Alkire, Roche and Vaz show that it is possible in that case to estimate upper and lower bounds on the extent to which the change in M_0 is due to 'movers' (changes in H) and the extent to which it is due to 'stayers' (changes in A). The same approach can be applied to estimate bounds on the extent to which the treatment effect $ATE_{c(k)}$ (6) is due to change in individuals' status and changes in intensity of missed objectives among those who fail to attain the overall goal of the program.

b) Decomposition of treatment effect by dimension

It is natural to ask to what extent our treatment effect of interest is driven by treatment effects on the different dimensions. The adjusted headcount ratio M_0 has the attractive property (Alkire and Foster 2011; Alkire et al 2015) that it may be decomposed by dimension as

$$M_0 = \sum_{d=1}^{D} w_d \, CH_d \tag{12}$$

where CH_d is the censored headcount ratio of indicator d,

$$CH_d = \frac{1}{\mathrm{N}} \sum_{n}^{\mathrm{N}} h_n I(x_{nd} < z_d). \tag{13}$$

In the present context the censored headcount ratio corresponds to the proportion of the sample who have both failed to achieve the overall goal and have missed the objective for that particular dimension. (Missed objectives among those who have achieved the overall goal are not counted.) Recalling that $A\hat{T}E_{c(k)} = M_0^T - M_0^C$, we can decompose both M_0^T and M_0^C by dimension and factorize to show that

$$A\hat{T}E_{c(k)} = \sum_{d=1}^{D} w_d \, A\hat{T}E_d \tag{14}$$

where $A\hat{T}E_d = CH_d^T - CH_d^C$, an estimate of the average treatment effect on the censored headcount ratio of indicator d. The decomposition of $A\hat{T}E_{c(k)}$ by dimension is simply the weighted sum of the average treatment effects on the censored headcount ratios. The decomposition of $A\hat{T}E_c$ by dimension is very similar, but using the uncensored rather than censored headcount ratios, $H_d = \frac{1}{\mathrm{N}} \sum_{n}^{\mathrm{N}} h_n$.

c) Decomposition of treatment effect by population sub-groups (heterogeneity analysis)

The adjusted headcount ratio M_0 also has the property of decomposability by population subgroups, provided the subgroups are mutually exclusive and collectively exhaustive. With $g = 1, \dots, G$ mutually exclusive and collectively exhaustive subgroups,

$$M_0 = \sum_{g=1}^{G} \frac{N_g}{N} M_0^g \tag{15}$$

where M_0^g is the within-subgroup adjusted headcount ratio, and N_g is the number of individuals in the subgroup. As above we can decompose both M_0^T and M_0^C by subgroup and factorize to show that

$$A\hat{T}E_{c(k)} = \sum_{g=1}^{G} \frac{N_g}{N} \, A\hat{T}E_g \tag{16}$$

where $\widehat{ATE}_g = M_0^{g,T} - M_0^{g,C}$, the estimate of the average treatment effect within subgroup g. The decomposition of $\widehat{ATE}_{c(k)}$ by subgroup is simply the population-weighted average of the average treatment effects within subgroups. The decomposition of $\widehat{ATE}_c$ by subgroups is exactly analogous, the population-weighted average of the average treatment effects within subgroups.

Subgroups of interest may be geographical or demographic categories; where baseline data is available, pre-intervention levels of the outcome indices may also provide a grouping of interest. In practice, it may be most natural to estimate the heterogeneous treatment effects by including interaction terms of treatment and group indicators in a regression model; as the model is more complex than a binary treatment regression and the dependent variables of interest are limited this should be taken into account in the specification.

d) Assessment of treatment effect on specific coupled deprivations

An attractive feature of the AF approach to multidimensional poverty measurement is that it is able to take account of the joint distribution of deprivations in different dimensions. A society in which deprivations in different dimensions are concentrated among the same people will be evaluated as more poor than a society in which deprivations in different dimensions are dispersed independently across different people.

We now consider how we may assess the effect of a program or policy intervention on specific coupled objectives. We may achieve this by estimating the effect of treatment on a categorical variable that distinguishes profiles representing different possible combinations of the multiple objectives. Thus, we suggest a multinomial logistic regression.

3. Building an index of program performance

In the development of the methodological framework above we took as given the target unit, the indicators for the multiple objectives, their weights and minimum achievement thresholds, as well as the cutoff k that represents the overall goal of the program. In any real-world evaluation task, these attributes and parameters of the outcome index must be selected by the evaluator.

We need to distinguish two scenarios. The first is when one is interested in measuring the impact of a policy intervention in reducing multidimensional poverty as defined by an official national MPI. In this case, the outcome of interest would be the existing national MPI, regardless of the initial objectives of the policy under analysis, so no further index design needs to be carried out.

The second scenario is when one is interested in examining the effect of a program or policy intervention on multiple specified objectives. In this case, the attributes and parameters of the outcome index should be specifically chosen to capture the objectives and expected outcomes of the policy. The rigor of this type of evaluation will depend on the quality of the multidimensional index used, so care must be exercised in the selection of the following characteristics.

(1) **The unit of identification,** n. The first question tends to be whether we want to identify the success of the program at the individual or household level. The answer will depend on the characteristics of the program or policy under analysis (target unit, objectives, etc.), as well as on the data available (or that can feasibly be collected).

(2) **The indicators,** x_d. These should reflect the expected outcomes of the program in the timeline of the evaluation. Although health might be an important dimension of poverty, if the program under analysis is not expected to have a direct or indirect effect on health, the index should not include indicators of health. Cash transfers conditional on children attending school are expected to increase the investment on children's health and education, which is expected to lead to higher earnings when these children became adults and enter the labor force. However, if we are measuring the impact of a cash-transfer in the short term, the level of earnings of children should not be included in the index.

(3) **The minimum achievement thresholds,** z_d. These should be based on the objectives of the program, and should allow space for improvement. If most children are attending school at least up to class 6, if we want to measure a policy impact on attendance we should use a threshold higher than class 6.

(4) **The indicators' weights**, w_d. These should reflect the relative importance of the different indicators/objectives. The selection of weights needs to take into consideration the relationship between indicators. For example, if two indicators have a significant overlap (people tend to be deprived in both simultaneously), a structure of equal weights across indicators might double count the same phenomenon.

(5) **The overall goal cutoff,** k. This parameter should be defined in line with the objectives of the program and allowing space for improvement.

4. An illustrative example

In this section we construct a very simple example to highlight the potential advantages of using the AF indices, rather than an array of disconnected indicators, as the outcome of interest in assessing the performance of a policy intervention with multiple objectives.

Consider a program that aims to ensure: (1) that individual's weekly income is at least 100 monetary units; (2) every adult has at least eight years of education; (3) no adult is malnourished, defined as not having a body mass index below 18.5; every adult lives in a household with (4) improved sanitation and (5) access to improved drinking water. Suppose that in order to assess the impact of the program, information regarding each indicator of achievement (income, years of education, body mass index, improved sanitation and improved drinking water) was collected for a group of individuals that benefited from the program – the treated group – and a comparable group of individuals who did not benefit from the program – the control group. For simplicity assume that the program was perfectly randomized, so the observable and unobservable characteristics, as well as the achievements, of control and treatment groups are identical at the baseline; and suppose that the achievements of the control group have not changed at all between the baseline and post-treatment survey. In these circumstances, the program's impact can be measured by simply comparing the outcomes of the treated group before and after the program. Table 1 displays the achievement matrices for the individuals in treated group before and after the program, as well as the vector of minimum achievements established by the program's goals.

Comparing the average achievement before and after the program we obtain the average impact of the program in each indicator. The results displayed in the bottom of Table 1 show that the program increased individual's average income by 6.25 monetary units; it raised the average number of years of education by 0.375 years; it increased the average body mass index by 0.5 units; and it improved the probability of having access to improved sanitation and drinking water by 25 percentage points. Although these results show that the program had a positive impact on all indicators, they do not show how close the program is to achieving its objectives. For instance, looking at the matrices of achievements we see that the increase in income has only benefited individuals who had already an income of 100 monetary units. The improvement in years of schooling has benefited two individuals who had an achievement below the minimum aimed by the program, but while one of the individuals achieved the program's objective, the other remained below. The program's impact on body mass index benefited the individuals who had the lowest achievement in that indicator, but the improvement was not enough to make

Table 2: Matrices of missed objectives before and after treatment, number of missed objectives and average impact on raw headcounts

	Matrices of missed objectives											
	Baseline						Post-treatment					
	Income	Edu	BMI	Sanit	Water	No. missed objectives	Income	Edu	BMI	Sanit	Water	No. missed objectives
1	1	1	1	1	1	5	1	1	1	0	0	3
2	1	1	1	1	1	5	1	0	1	0	0	2
3	1	1	1	0	0	3	1	1	1	0	0	3
4	1	1	1	0	0	3	1	1	1	0	0	3
5	0	1	1	0	0	2	0	1	1	0	0	2
6	0	1	1	0	0	2	0	1	1	0	0	2
7	0	1	0	0	0	1	0	1	0	0	0	1
8	0	0	0	0	0	0	0	0	0	0	0	0
	Program's average impact on raw headcounts											
	Raw headcount at baseline						0.50	0.88	0.75	0.25	0.25	2.625
	Raw headcount at post-treatment						0.50	0.75	0.75	0.00	0.00	2.000
	Program's average impact						0.00	-0.13	0.00	-0.25	-0.25	-0.625

these people achieve the program's objective. Therefore, the assessment of the program's impact based on the level of achievements, without reference to the minimum thresholds aimed at, provides an incomplete picture of the program's performance.

The method proposed here focuses on the missed objectives, i.e. on the deprivations experienced by the individuals, instead of on their levels of achievement. One of the first steps in the computation of an AF index is to reduce the information of the achievement matrix and the vector of minimum achievement thresholds into a matrix of deprivations (or missed objectives). The matrices of missed objectives before and after the program are displayed in Table 2. Missed objectives are identified with the value one, a zero means that the individual has achieved the objective in that specific indicator. Summing the missed objectives across columns we obtain the number of missed objectives for each individual. For instance, at the baseline individuals 1 and 2 did not achieve the minimum threshold in any of the five indicators, while individual 8 had attained the threshold in all indicators.

As an intermediate step we can look at the impact on the raw missed objective headcounts. This will show the extent to which the program is leading to attainment of each objective separately. The results displayed at the bottom of Table 2 show that the program has made no progress in terms of ensuring minimum income and nourishment. We have seen that in both dimensions the program has actually generated an increase in the respective average achievement (income and body mass index). However, if we focus on the program's goals, the achievement of minimum thresholds in each dimension, those gains are not relevant. In the case of income, the increase has only benefited individuals who already earned a sufficient income. In the case of body mass index, although the improvements benefited people who were malnourished, those improvements were not sufficient to raise the individuals' achievements above the minimum threshold. Thus, for neither of these dimensions was the program successful in achieving the respective goal, although in the case of nutrition it may be on track to do it.

We can also look at the program's impact in reducing the average number of missed objectives of individuals. In this example, the program reduced the average number of individuals' missed objectives from 2.625 to 2.

As some missed objectives may reflect individuals' preferences rather than lack of capabilities (for instance, a fashion model who chooses to maintain a body mass index below 18.5), the program's overall goal might be to ensure that each individual does not have K or more missed objectives[8]. Thus, when assessing the impact of the program, besides focusing on missed objectives, we also focus on the individuals who are missing K or more objectives. The AF index adjusted headcount (M_0) and its partial indices incidence (H) and intensity (A) allows us to do precisely that.

Table 3: Program's multidimensional impact

	Overall goal: no one misses… goals or more				
	K =1	K = 2	K = 3	K = 4	K = 5
Levels					
Baseline					
Incidence	0.88	0.75	0.50	0.25	0.25
Intensity	0.60	0.67	0.80	1.00	1.00
Adjusted headcount ratio	0.53	0.50	0.40	0.25	0.25
Post-treatment					
Incidence	0.88	0.75	0.38	0.00	0.00
Intensity	0.46	0.50	0.60	0.00	0.00
Adjusted headcount ratio	0.40	0.38	0.23	0.00	0.00
Program's impact					
Incidence (change)	0.00	0.00	-0.13	-0.25	-0.25
Intensity (change)	-0.14	-0.17	-0.20	-1.00	-1.00
Adjusted headcount (change)	**-0.13**	**-0.13**	**-0.18**	**-0.25**	**-0.25**

Table 3 presents the AF indices for treated individuals before and after the program, for different K values.[9] Before proceeding, we discuss how to interpret these figures. Suppose the program's overall goal is to ensure that all individuals don't miss three or more objectives, $K = 3$. The incidence is the percentage of individuals that do not have the minimum achievements in at least three of the indicators; this percentage reduced from 50 to 37.5 percent (because individual 2 who missed five objectives at the baseline, is only missing two after the program). The intensity is the average number of missed objectives for those who are missing at least three. This percentage reduced from 80 to 60 percent (mostly because individual 1's number of missed objectives reduced from five to three)[10]. The adjusted headcount, as the name suggests, is the headcount of people missing at least three objectives adjusted by the average number of objectives they are missing; an higher adjusted headcount means that the program is further

8 Here K simply counts objectives, so may be between 0 and 5. It is closely related to, but not the same as k in sections 2 and 3 above, which is a weighted sum of objectives, between 0 and 1.

9 In this example we assumed that all of the program's goals have equal weights.

10 Changes in intensity might be driven by changes in deprivations among those who remained poor or missing the overall goal, but also by changes in incidence.

from achieving its overall goal. The adjusted headcount reduced from 0.400 at the baseline to 0.225 after the treatment. Selection of higher K values means that the overall goal of the program is to focus on people who miss a higher number of objectives (have a higher number of deprivations).

Table 3 shows that the program reduced the adjusted headcount for all possible overall goals, i.e. for all possible K values. Furthermore, it shows that the program's impact was larger for higher values of K. This means that the program had a larger effect on individuals who experience higher number of missed objectives / deprivations. The use of an AF index to assess the program's impact enables us to examine if the program has a differentiated impact for people with different levels of missed objectives / deprivations. This information might be very relevant for policy purposes as it might help us understand how the program works.

5. An Empirical Application: Multidimensional Impact Evaluation of *Progresa*

The main purpose of this section is to illustrate how the AF methodology can be used in impact evaluation, as well as to demonstrate the set of analyses that can be done within that framework.

5.1 Program and data

Mexico's social assistance program Progresa (renamed Oportunidades in 2002) is a landmark in poverty reduction policy design. This program aimed at alleviating poverty by using conditional cash transfers to induce families to invest in education, health and nutrition (Skoufias, 2005). Families identified as poor were given (1) educational grants (which value depended on children's grade and gender) conditional on children attending at least 85% of the classes; (2) monetary transfers and nutritional supplements conditional on regular health care visits and attendance of health talks. The average monthly transfer ascended to 20 percent of families' monthly expenditure before the program (Skoufias, 2005).

Unlike most previous programs, Progresa was committed to a rigorous evaluation since its inception. First, the sequential expansion of the program was used to adopt an experimental design in the first years of the program's operation. This was one of the first applications of experimental methods to a large-scale development program. Second, data was collected from households in areas that were already eligible to the program and households in areas that were not yet, before and after the implementation of the program. This effort resulted in the construction of a panel dataset covering approximately 24,000 households from 506 localities (320 eligible to the program and 186 selected as control), in seven states (Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz). The baseline was constructed using data from the Households' Socio-economic Characteristics Survey (Encuesta de Caracteristicas Socioeconomicas de los Hogares, ENCASEH), conducted in 1997, and the Households Evaluation Survey (Encuesta de Evaluacion de los Hogares, ENCEL) conducted in March of 1998. Then, there were three follow-up Household Evaluation Surveys conducted in October of 1998, March of 1999 and November of 1999. This data is publicly available at www.evaluacion.oportunidades.gob.mx:8010. In this study, we will focus on the sample of individuals residing in households that were identified as 'poor' according to the program's criteria and, thus, were eligible to benefit from Progresa. Information on the sample size and attrition rates can be found in Table A.1 in Appendix.[11]

The evaluation of *Progresa* was conducted by International Food Policy Research Institute (IFPRI). The program's short-term impact in single indicators is very well documented. In the area of children's education, the program had a positive impact on school enrolment, but no impact on attendance (Schultz, 2000); it increased the number of grades completed (Behram, Parker and Todd, 2005); it decreased the school grade gaps (Behram, Sengupta and Todd, 2000, 2005); and it reduced child labor (Parker and Skoufias, 2000). Regarding health, *Progresa* reduced the probability of illness among children under five years old and increased the number of visits to public health centers (Gertler, 2000). In terms of nutrition, the program increased households' expenditure on food (Hoddinott and Skoufias, 2004).

In the next sub-sections we present an evaluation of the short-term impact of *Progresa* using a multidimensional outcome index as outlined in section 2 above.

5.2 Constructing a multidimensional index of Progresa's performance

Given that our purpose is mainly illustrative, we build a simple multidimensional index using a small set of indicators. We selected the indicators based on the evaluation literature. The indicators' minimum achievement thresholds reflect the program's minimum goals. The structure of our multidimensional index is summarized in Table 4.

11 The attrition rates are relatively high. At the level of the household the survey attrition rate is around 6% after two waves. However, if we also consider the sample drop due to missing information on the indicators we are using, the attrition increases to around 16%. Nonetheless, as the attrition rates are similar across control and treatment groups, we hope they do not compromise the comparability of the two groups.

Table 4: Indicators and weights

Indicator	Deprived if:	Weight
School attendance	at least one member aged 6-14 attended less than 90% of the school days (past month) OR is not enrolled	0.2
No child labor	at least one member aged 8-14 had a job or worked during last week (even if unpaid)	0.2
Children's health	at least one member aged 0-2 was ill in the past 4 weeks for more than 5 days	0.2
Health visits for nutrition monitoring	at least one member aged 0-2 has not made any visit in the past 6 months (1)	0.2
Assets	at least one of the following assets: refrigerator, television or radio	0.2

(1) The question regarding the health visits for nutrition monitoring in the survey ENCEL 1998 March referred to the past 12 months. Therefore, this indicator has a reference period larger at the baseline than in the subsequent periods.

Unfortunately, there are no good indicators to assess the short-run impact of the program on nutrition. There is information about children's weight and height only in 2003 and 2007, and there is no baseline information about food consumption. Thus, we could not include any indicator that directly measured progress in that area.

As the target unit of the program was the household, the unit of identification for the analysis will also be the household. Nevertheless, our computations will be made in terms of individuals; this means that all individuals living in a household that has not met the threshold in school enrolment will be considered below the threshold in that indicator, regardless of their age or specific situation.[12] As all indicators are defined with reference to children, the status of the household is dependent on its demographic structure.

In this application, we will assume that the overall goal of the program is to achieve minimum thresholds in all dimensions. Thus, an individual will be considered not to have met the goal if he or she lives in a household that has not met the threshold in at least one of the five indicators, $k = 0.2$.

5.3 Baseline balance tests

Table 5 shows the balance tests on the baseline data for the percentage of people deprived in each indicator, the distribution of the deprivation score and M_0, H and A for the overall goal threshold 20%. There are practically no statistically significant differences; the only exception is the proportion of people deprived in all indicators.

12 We could also make all computations in terms of households. However, as there is the possibility that poor households are on average larger than non-poor households, we find that reporting percentages of people is more informative than reporting percentages of households.

Table 5: Differences between treated and control groups at the baseline

	Control	Treated	Difference	Standard error
Indicators				
Attendance	24.5%	22.6%	-0.019	(0.015)
No child labor	10.2%	11.5%	0.013	(0.012)
Children's health	8.7%	9.9%	0.012	(0.009)
Health visits for nutrition monitoring	8.6%	10.3%	0.016	(0.013)
Assets	33.4%	36.2%	0.028	(0.024)
Weighted sum of missed objectives				
0%	39.6%	38.2%	-0.015	(0.020)
20%	40.2%	39.6%	-0.006	(0.013)
40%	15.9%	16.9%	0.010	(0.012)
60%	3.8%	4.5%	0.007	(0.006)
80%	0.6%	0.9%	0.002	(0.002)
100%	0.0%	0.1%	0.001*	(0.001)
Outcome indices				
Incidence	60.4%	61.8%	0.015	(0.020)
Intensity	28.3%	29.3%	0.010	(0.006)
Adjusted headcount ratio	0.171	0.181	0.010	(0.008)

Note: Standard errors reported; these are corrected by clustering at the location (enumeration area) level.

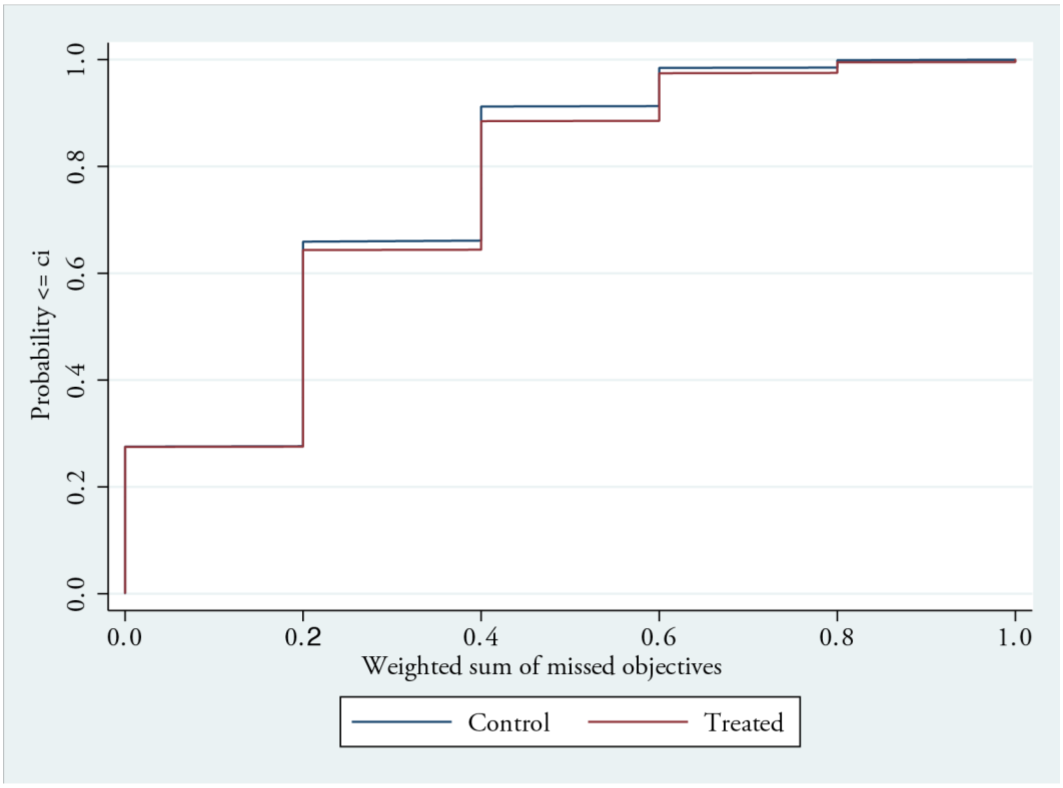
Figure 1 shows that the cumulative distribution of the weighted sum of missed objectives in targeted and control areas at baseline seems to be very close. This suggests that regardless of the overall goal threshold we select, the incidence and intensity will be very similar in treated and control areas. The Kolmogorov-Smirnov test does reject the null hypothesis of identical distributions of the weighted sum of missed objectives, with D= 0.0218 and corresponding p-value of 0.000.[13] However, at baseline the empirical distribution for the treated group lies to the right of the empirical distribution for the control group, corresponding to higher weighted sum of missed objectives. If there were any persistent effect of this imbalance it would tend to attenuate our treatment effect estimates.

The percentage of people living in households that are not missing any objective is 39.6% in control areas and 38.2% in treated areas. These percentages seem very high considering we are looking only at the sample of eligible households. A reason for this is that the households' selection into the program was based on a discriminant linear analysis including household

13 The test is conservative with discrete distributions, so we may be confident in the rejection.

income and dwelling characteristics, while the indicators included in our multidimensional index are mostly related to children and the goals of the program.

Figure 1: Cumulative distribution of the weighted sum of missed objectives at baseline



Source: Authors' calculations.

5.4 Average treatment effects

We adopt standard regression specifications to estimate the average impact of randomized experiments. Let Y_i, represent the outcome variables for individual *i*. First, we consider a specification that uses only post-treatment data:

$$Y_i = \alpha + \beta P_i + \varepsilon_i$$

Where P_i is an indicator variable for a community targeted by *Progresa*, and β is the cross-section estimate of the program's effect on the eligible individuals living in treated locations. The second specification also uses the baseline data:

$$Y_{i,t} = \alpha + \beta_1 P_i + \beta_2 T1_{i,t} + \beta_3 T_{i,t} P_i + \varepsilon_{i,t}$$

Where $T1_{i,t}$ is an indicator of the first post-treatment wave, and β_3 is the difference-in-difference estimate of the program's effect on the eligible individuals living in treated locations. Given the availability of baseline data, the difference-in-difference specification may yield more precise estimates of the treatment effects; as the baseline balance tests give us no reason to doubt the randomization of treatment, the difference-in-difference approach is not necessary for identification.

As discussed in the introduction, standard practice to evaluate the impact of programs with multiple objectives has been to measure and assess the statistical significance of differences in a dashboard of single indicators across treated and control groups. For each of the indicators selected, Table 6 displays the treated and control groups' means in each time period and the cross-section and difference-in-difference estimates of impact. Based on the difference-in-difference estimates, after one period, the program significantly reduces the percentage of people not in school attendance, no child labor and health care visits. After two periods the program has a significant effect on all indicators except school attendance. These results, although valuable, do not inform us about the program's impact weighted sum of missed objectives experienced by people. The program might be judged differently if it affects mostly people who have a higher number of missed objectives to start with, instead of people who have missed just one objective.

Table 6: Average impact on single indicators

Objectives / Indicators	% of individuals living in households not meeting threshold in…			Program's effect (p.p.)			
	Time	Control areas	Treated areas	Dif.		Dif.-in-Dif.	
				Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
School attendance	0	24.5%	22.6%				
	1	24.0%	16.4%	-7.6***	(0.015)	-5.7***	(0.015)
	2	25.7%	23.2%	-2.5	(0.025)	-0.6	(0.021)
No child labor	0	10.2%	11.5%				
	1	7.0%	5.3%	-1.8**	(0.007)	-3.1**	(0.012)
	2	6.4%	5.8%	-0.6	(0.009)	-1.9*	(0.012)
Children's health	0	8.7%	9.9%				
	1	6.1%	5.9%	-0.3	(0.006)	-1.4	(0.009)
	2	5.2%	4.6%	-0.6	(0.006)	-1.8*	(0.010)
Health visits	0	8.6%	10.3%				
	1	9.8%	6.0%	-3.8***	(0.013)	-5.5***	(0.013)
	2	6.0%	3.0%	-3.0***	(0.011)	-4.7***	(0.013)
Assets	0	33.4%	36.2%				
	1	43.5%	44.6%	1.1	(0.028)	-1.8	(0.020)
	2	39.0%	37.6%	-1.3	(0.028)	-4.2*	(0.022)

Note: Standard errors are corrected by clustering at the locality level.
* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7 presents the average effect on the weighted sum of missed objectives, incidence, intensity and adjusted headcount ratio. For each dependent variable we present the cross-section as well as the difference-in-difference estimates of impact. The average impact of the program in these variables is statistically significant in both periods. The weighted sum of missed objectives reduced by more than 2 p.p. in both periods. Considering the overall goal threshold of 20%, we see that the program has both increased the proportion of people achieving the overall goal (incidence), and reduced the weighted sum of missed objectives among those who do not (intensity). The estimates of the effect for period 1 and period 2 are not statistically different, while the point estimates for period 2 are always lower than the point estimates for period 1. This suggests that the multidimensional impact of *Progresa* did not increase with the duration of the intervention. A possible explanation is that the majority of the impact occurs due to the program's conditionality. The fact that the indicators that registered a larger effect were those associated with the conditions of the program, school attendance and health care visits, seems to corroborate this hypothesis.

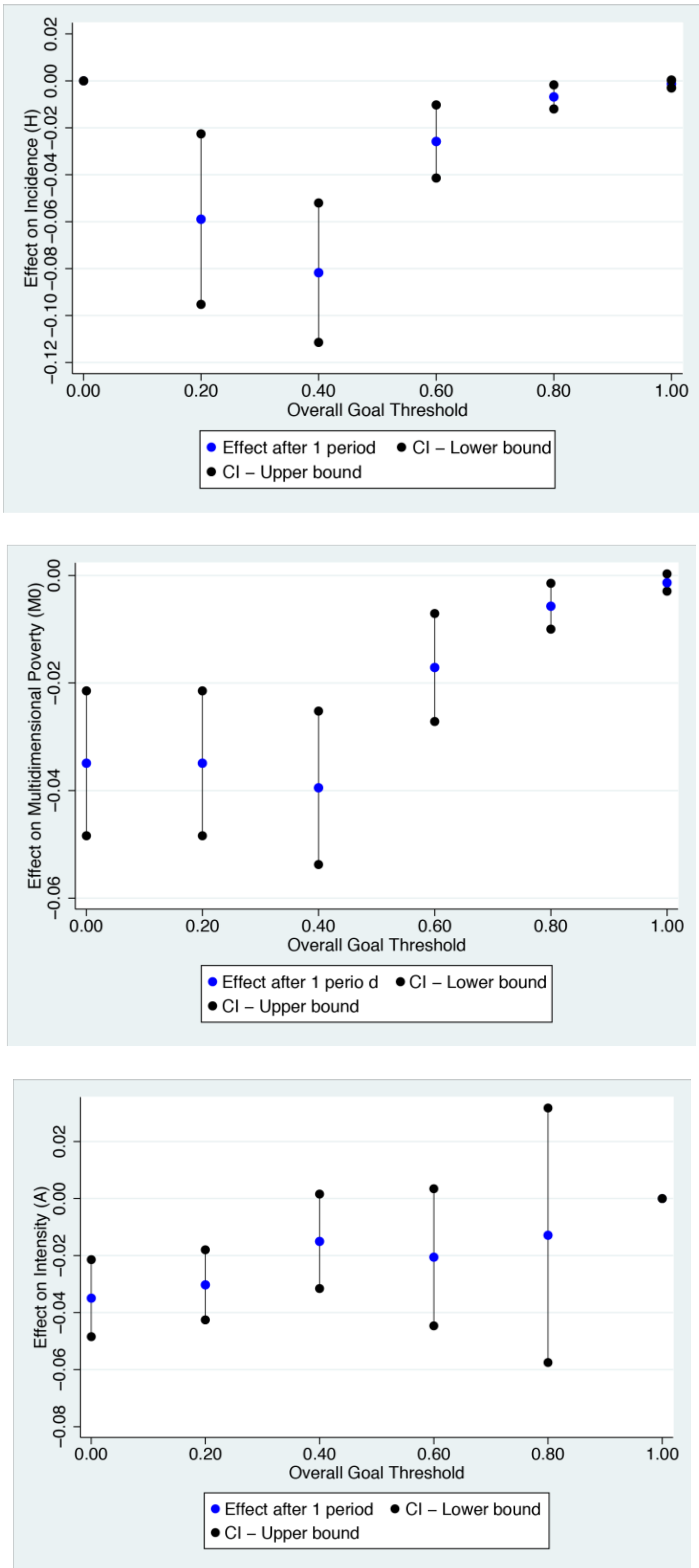
Table 7: Average impact on multidimensional indices

Multidimensional indices	Time	Control areas	Treated areas	Program's effect (p.p.)			
				Dif.		Dif.-in-Dif.	
				Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Weighted sum of missed objectives	0	17.1%	18.1%				
	1	18.1%	15.6%	-2.5***	(0.007)	-3.5***	(0.007)
	2	16.4%	14.8%	-1.6*	(0.009)	-2.6***	(0.007)
Incidence (k=0.2)	0	60.4%	61.8%				
	1	64.4%	60.0%	-4.4**	(0.021)	-5.9***	(0.019)
	2	59.9%	56.5%	-3.4	(0.023)	-4.9***	(0.019)
Intensity (k=0.2)	0	28.3%	29.3%				
	1	28.1%	26.0%	-2.1***	(0.005)	-3.0***	(0.006)
	2	27.4%	26.3%	-1.2	(0.008)	-2.1***	(0.008)
Adjusted headcount ratio (k=2)	0	0.171	0.181				
	1	0.181	0.156	-0.025***	(0.007)	-0.035***	(0.007)
	2	0.164	0.148	-0.016***	(0.009)	-0.026***	(0.007)

Note: Estimates regarding the program's effect on intensity are based on regressions considering only the sample of individuals identified as poor. Standard errors are corrected by clustering at the locality level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 2: Average effect on outcome indices across different overall goal thresholds



Source: Authors' calculations.

How would these results change if we had considered a different overall goal threshold? Figure 2 depicts the average effect of the program in period 1 (difference-in-difference estimate) on incidence, intensity and adjusted headcount ratio (censored weighted sum of missed objectives) for all possible thresholds. This graph shows that the effect of the program on incidence and on the adjusted headcount ratio is significant for all thresholds below 1. This suggests that people with higher weighted sum of missed objectives, who are the only ones to be identified as not meeting the overall goal when we use higher thresholds, are benefiting from the program to some extent.

5.5 Heterogeneous effects

5.5.1 Across categories of the weighted sum of missed objectives

Table 8 presents the estimates of an ordered logistic model in which categories of the weighted sum of missed objectives are regressed on the treatment. It displays the ordered log-odds estimate for being treated on the weighted sum of missed objectives, as well as the marginal effects of the program on each of the weighted sum of missed objectives categories. If a person is a beneficiary of *Progresa*, we find that his or her ordered log-odds of having higher weighted sum of missed objectives decreases by 26.2% (ceteris paribus). The marginal effect can be thought of as the change in probability of being in a particular category subject to receiving the treatment. We find a positive and significant marginal effect of the program on attainment of all objectives (weighted sum of missed objectives = 0) and a reduction in the probabilities for the remaining categories. The effect is most pronounced (in magnitude) for weighted sums of 0.2 and 0.4.

Table 8: Marginal effects of Progresa on weighted sum of missed objectives levels, ordered logit model

	Log-Odds	Std. Error
Beneficiary of *Progresa*	-0.262	0.086
Weighted sum of missed objectives	**Marginal Effects**	**Std. Error**
0	0.06125	0.01992
0.2	-0.02321	0.00735
0.4	-0.02941	0.00985
0.6	-0.00757	0.00263
0.8	-0.00098	0.00041
1	-0.00008	0.00005

5.5.2 Across different profiles of missed objectives

We disaggregated the weighted sum of missed objectives into finer categories that distinguish the different profiles of missed objectives that constitute a certain weighted sum of missed objectives. The profile categories with less than 100 observations were grouped together under "other profiles" due to the difficulty in calculating standard errors for such small bins. Table 9 presents the estimates of the multinomial logit model that regresses the profile categories on the treatment, considering as base the category 'attaining all objectives'.

Each line corresponds to a different profile of missed objectives. The X's identify the missed objectives in each particular profile. For instance, the first line corresponds to the profile in which the objective is missed only for the assets indicator. The shaded rows correspond to the coefficients that are statistically significant at least at 5% level.

Table 9: Estimates of *Progresa* impact on profiles of missed objectives, multinomial logit

Weighted sum of missed objectives	No. obs	School attendance	Child labor	Children's health	Health visits	Assets	Coefficient	P-Value
0	28,594	Base Category						
0.2	23,139					X	0.030	0.797
0.2	1,870				X		-0.613	0.001
0.2	1,716			X			0.183	0.191
0.2	685		X				0.176	0.526
0.2	5,789	X					-0.414	0
0.4	1,885				X	X	-0.412	0.134
0.4	1,270			X		X	-0.013	0.949
0.4	177			X	X		-1.068	0.038
0.4	674		X			X	-0.027	0.923
0.4	3,848	X				X	-0.490	0
0.4	455	X			X		-0.677	0.034
0.4	494	X		X			-0.417	0.163
0.4	1,537	X	X				-0.712	0
0.6	228			X	X	X	-0.637	0.24
0.6	433	X			X	X	-0.308	0.407
0.6	189	X		X		X	-0.418	0.318
0.6	981	X	X			X	-0.272	0.212
0.6	151	X	X		X		-1.710	0.007
0.6	105	X	X	X			-0.639	0.337
0.8	140	X	X		X	X	-0.937	0.133
	312	Other Combinations (mean weighted sum of missed objectives: 0.696)					-1.009	0.014

These estimates show that being part of the treatment group reduces the relative log odds of just a few profiles of missed objectives as opposed to the base category of no missed objectives at all, namely: missing the objective in health visits alone; school attendance alone; illness and health visits; school attendance and assets; school attendance and health visits; school attendance and child labor; school attendance, child labor and children health. The other profiles of missed objectives showed no significant decrease in likelihood of being falling into that profile, as a result of treatment. All these categories for which we obtained a statistically significant effect include at least one of the indicators related to the conditionality of the program, school attendance and health visits. This evidence suggests that the effect of the program on our outcome of interest is being driven mostly by the conditionality of the program.

5.5.3 Across baseline characteristics

We tested whether the program had heterogeneous effects on the weighted sum of missed objectives depending on the extent to which the individual had failed to achieve the objectives at baseline, adopting the specification

$$c_{i,t} = \alpha + \beta_1 P_i + \beta_2 c_{i,0} + \beta_3 c_{i,0} P_i + \varepsilon_{i,t}.$$

OLS estimates are presented in Table 10.

Table 10: Heterogeneity of treatment effect by baseline weighted sum of missed objectives (OLS Estimates)

	(1) Time=1	(2) Time=1	(3) Time=1	(4) Time=2	(5) Time=2	(6) Time=2
Treatment	-0.0350***	-0.0272***	-0.0241***	-0.0274***	-0.0191***	-0.0130*
	(0.00647)	(0.00574)	(0.00654)	(0.00697)	(0.00722)	(0.00681)
Baseline weighted sum of missed objectives		-0.726***	-0.714***		-0.742***	-0.718***
		(0.0136)	(0.0221)		(0.0135)	(0.0271)
Interaction of treatment and baseline weighted sum of missed objectives			-0.0185			-0.0366
			(0.0280)			(0.0309)
Constant	0.00899*	0.130***	0.128***	-0.00677	0.115***	0.111***
	(0.00490)	(0.00494)	(0.00531)	(0.00529)	(0.00525)	(0.00518)
N	61985	61985	61985	54740	54740	54740

Cluster robust standard errors are reported in parentheses
* p<0.1, ** p<0.05, *** p<0.01

As the dependent variable is constrained to lie between 0 and 1, it is appropriate to use a specification that reflects this constraint. Tobit results for the same model are presented in Table 11; all results are robust to the alternative specification.

There is no evidence for heterogeneity of the treatment effect according to the baseline weighted sum of missed objectives. This suggests that the impact of the program did not differ substantially across individuals who only marginally failed to achieve the objectives at baseline, and those who had failed to achieve most of the program's objectives.

Table 11: Heterogeneity of treatment effect by baseline weighted sum of missed objectives (Tobit Estimates)

	(1) Time=1	(2) Time=1	(3) Time=1	(4) Time=2	(5) Time=2	(6) Time=2
Treatment	-0.0352***	-0.0398***	-0.0389***	-0.0249*	-0.0297**	-0.0233*
	(0.0111)	(0.00929)	(0.0119)	(0.0140)	(0.0120)	(0.0139)
Baseline weighted sum of missed objectives		0.421***	0.424***		0.424***	0.447***
		(0.0194)	(0.0309)		(0.0212)	(0.0394)
Interaction of treatment and baseline weighted sum of missed objectives			-0.00517			-0.0359
			(0.0389)			(0.0446)
Constant	0.118***	0.0491***	0.0485***	0.0841***	0.0157	0.0117
	(0.00896)	(0.00866)	(0.00979)	(0.0106)	(0.00990)	(0.0111)
Standard error	0.241***	0.228***	0.228***	0.252***	0.240***	0.240***
	(0.00410)	(0.00375)	(0.00375)	(0.00454)	(0.00400)	(0.00399)
N	61985	61985	61985	54740	54740	54740

Cluster robust standard errors are reported in parentheses

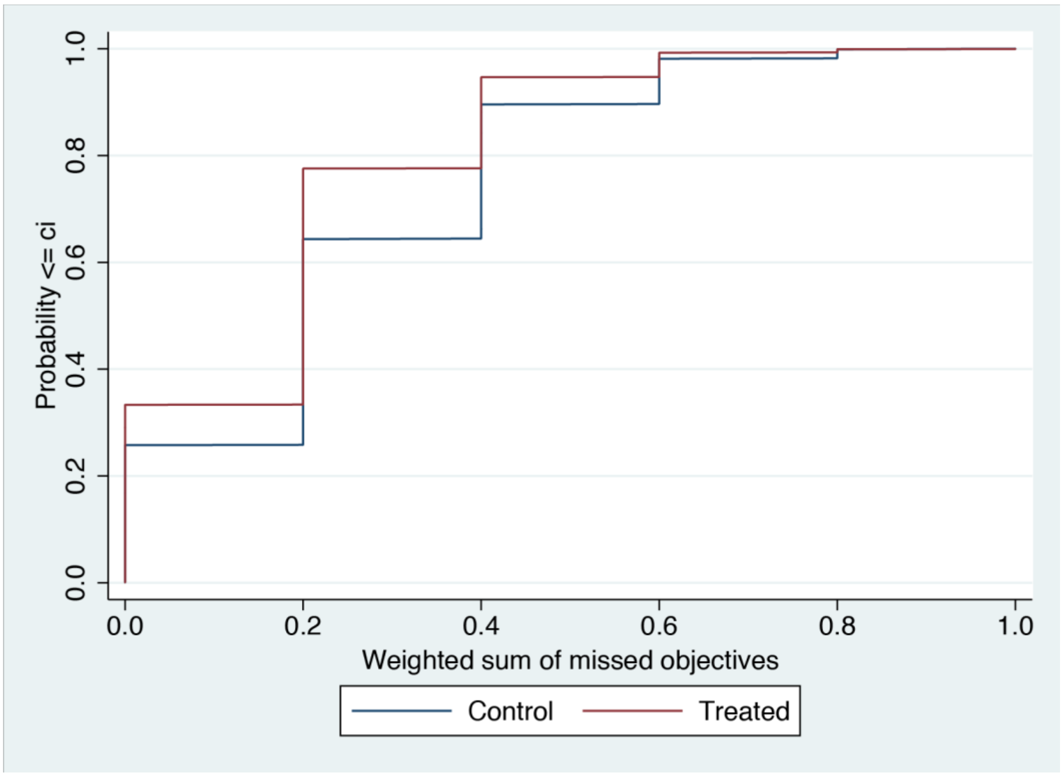
* $p<0.1$, ** $p<0.05$, *** $p<0.01$

There is no evidence for heterogeneity of the treatment effect according to the baseline weighted sum of missed objectives. This suggests that the impact of the program did not differ substantially across individuals who only marginally failed to achieve the objectives at baseline, and those who had failed to achieve most of the program's objectives.

5.6 Treatment effect on the distribution of the weighted sum of missed objectives

Another way of examining the impact of *Progresa* is to compare the cumulative distribution of the weighted sum of missed objectives among treated and control groups after the treatment. Figure 3 shows that the proportion of individuals in the lower end of the distribution (with fewer missed deprivations) is higher among the treated than among the control group. The Kolmogorov-Smirnov test strongly rejects the null hypothesis of identical distributions of the weighted sum of missed objectives at endline (period 1), with D = 0.0597 and corresponding p-value of 0.000.[14] This clearly demonstrates that *Progresa* has reduced the number of missed objectives among its beneficiaries.

Figure 3: Cumulative distribution of the weighted sum of missed objectives in period 1



Source: Authors' calculations.

6. Concluding remarks

This paper demonstrates how Alkire-Foster multidimensional indices can be adapted to evaluate the impact of policy interventions with multiple objectives, and suggests a set of additional analyses that can contribute to better understand the effect of such policies on coupled (or joint) deprivations. We use the conditional cash-transfer *Progresa* in Mexico to illustrate the proposed framework. We build a multidimensional index that reflects the objectives of the program, and examine the impact of the program with reference to that index. The analyses suggest that the effect on our multidimensional index is driven mainly by the conditions of the program on school attendance and health visits.

In our empirical application, the risk of selection bias was relatively low because *Progresa*, in its early stages, was randomly assigned at the community level. Further research is required to examine how this framework can be used in combination with non-experimental evaluation techniques such as propensity-score matching and instrumental variables.

14 The test is conservative with discrete distributions, so we may be confident in the rejection.

References

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

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G. and Vaz, A. (2013). 'The women's empowerment in agriculture index', *World Development*, *52*, 71–91.

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

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

Alkire, S. and Seth, S. (2013). 'Selecting a targeting method to identify BPL households in India', *Social indicators research*, *112*(2), 417–446.

Anderson, M. L. (2008). 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects', *Journal of the American statistical Association*, *103*(484), 1481–1495.

Azevedo, V. and Robles, M. (2013). 'Multidimensional targeting: Identifying beneficiaries of conditional cash transfer programs', *Social Indicators Research*, *112*(2), 447–475.

Behrman, J. R., Sengupta, P. and Todd, P. (2000). 'The impact of PROGRESA on achievement test scores in the first year', *Washington DC: International Food Policy Research Institute. Mimeo*.

Behrman, J. R., Sengupta, P. and Todd, P. (2005). 'Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico', *Economic development and cultural change*, *54*(1), 237–275.

Bouillon, C. P. and Yáñez-Pagans, P. (2011). 'Dynamic consistency of multidimensional and income targeting: An application for Mexico using panel data information', *IDB Workin Paper Series No. IDB-WP-250*, Inter-American Development Bank.

Gertler, P. (2000). *Final report: The impact of PROGRESA on health*. International Food Policy Research Institute, Washington*, DC*, 2000.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. and Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

Hoddinott, J. and Skoufias, E. (2004). 'The impact of PROGRESA on food consumption', *Economic development and cultural change*, *53*(1), 37–61.

Loschmann, C., Parsons, C. R. and Siegel, M. (2015). 'Does shelter assistance reduce poverty in Afghanistan?', *World Development*, *74*, 305322.

McKenzie, D. (2012). 'Beyond baseline and follow-up: The case for more T in experiments', *Journal of Development Economics*, *99*(2), 210–221.

O'Brien, P. C. (1984). 'Procedures for comparing samples with multiple endpoints', *Biometrics*, 1079–1087.

Parker, S. and Skoufias, E. (2000). *The impact of PROGRESA on work, leisure and time allocation*. Washington, DC: International Food Policy Research Institute.

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

Robano, V. and Smith, S. C. (2014). 'Multidimensional targeting and evaluation: A general framework with an application to a poverty program in Bangladesh', *OPHI Working Paper* 65, University of Oxford.

Rubin, D. B. (2005). 'Causal inference using potential outcomes: Design, modeling, decisions', *Journal of the American Statistical Association*, *100*(469), 322–331.

Salazar, A., Cuervo, Y. D. and Pinzón, R. P. (2011). 'Índice de pobreza multidimensional para Colombia (IPM-Colombia) 1997-2010', *Archivos de Economía*, (382).

Schultz, T. P. (2000). *Impact of PROGRESA on school attendance rates in the sampled population*. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, DC.

Skoufias, E. (2005). *PROGRESA and its impacts on the welfare of rural households in Mexico* (Vol. 139). International Food Policy Research Institute.

Appendix: Sample Sizes and Attrition

Table A.1: Sample size and attrition

Datasets	Sample of eligible households with information for all indicators (1)								
	Sample size					Atrition rates (%)			
	Control areas		Treatment areas		All	Control areas		Treatment areas	
	Households	Individuals	Households	Individuals	Individuals	Households	Individuals	Households	Individuals
ENCASEH 97 + ENCEL 98 March	4,319	27,106	7,220	45,014	72,120				
ENCEL 98, October	4,727	28,137	7,928	46,535	74,672				
Panel with two time periods (2)	**3,617**	**22,479**	**6,109**	**37,540**	**60,019**	**16.25**	**17.07**	**15.39**	**16.60**
ENCEL 99, March	4,396	25,782	7,368	42,684	68,466				
Panel with three time periods	**3,125**	**19,424**	**5,311**	**32,801**	**52,225**	**27.65**	**28.34**	**26.44**	**27.13**
ENCEL 99, November									
Panel with four time periods									

(1) Based on the original eligibility criterion, 'pobre' and considering only resident household members.