Sensitivity matters. Comparing the use of multiple indicators and of a multidimensional poverty index in the evaluation of a poverty eradication program.

August 6, 2020

Abstract

This paper investigates the performance of the UNDP multidimensional poverty index (global MPI) and of Millennium Development Goals (MDGs) indicators in the impact evaluation of the SADA-Northern Ghana Millennium Village Project (MVP). We find that the project had a limited impact on MDG indicators and yet a large impact on the global MPI. We assess the robustness of the impact of MVP on the global MPI and we find that this was largely driven by changes in few MDG indicators. We conclude that the MVP had a limited impact on living standards, and that the global MPI should be used with caution in the evaluation of development programmes.

1 Introduction

In this paper we compare the performance of two different ways of measuring outcome indicators in the evaluation of poverty eradication programmes. Poverty is multidimensional and poverty eradication programmes affect multiple outcomes at the same time. A common approach to the evaluation of these programmes employs a "dashboard" of indicators (Ravallion, 2011), and tracks progress on each outcome separately. A second popular approach collapses many indicators in a single index, and measures the impact of interventions on the index.

Both methods have advantages and disadvantages. The analysis of multiple indicators produces results that are easy to interpret. The use of multiple indicators allows identifying the areas in which a programme was successful, and helps mapping impacts from activities to outputs, and to final outcomes. On the other hand, the use of multiple indicators can produce mixed and inconclusive results, as the researchers struggle to make sense of different outcomes in different dimensions of well-being. In addition, the use of multiple indicators is susceptible to the multiple comparisons problem stressed in the statistical literature (Efron and Hastie, 2016), whereby, when testing many hypotheses at the same time, there is a chance of finding positive impacts even if the project had no impact at all.

The second method reduces various poverty dimensions to a single indicator - a poverty index - and in this way greatly simplifies the task of assessing the overall impact of an intervention. In addition, by construction, the reduction in the number of dimensions eliminates the statistical problem of testing multiple hypotheses. Poverty indices, however, can be built in different ways, and their construction requires a number of seemingly arbitrary choices about their constitutive elements and relative weights. Impacts on indices are also difficult to interpret, particularly when they are built ad-hoc, and cannot be compared to those of studies using different indices.

In this paper we compare these two methods in the impact assessment of the SADA[1]-

[1]The Savannah Accelerated Development Authority (SADA) was an agency of the Government of Ghana

Northern Ghana Millennium Village Project (MVP) implemented between 2012 and 2016. The SADA-MVP was one of 15 Millennium Villages established in Sub-Saharan Africa by the Millennium Promise, the Earth Institute at Columbia University and the United Nations Development Program with the goal of achieving the MDGs at the village-level (Sanchez et al., 2007). The project design followed the recommendations of the UN Millennium Village project, which set out the ambitious goal of ending African poverty. The MDGs were identified not as a final goal, but as an intermediate target towards the goal of breaking the poverty trap (Sachs et al., 2004), because they reflected the fact that poverty is multidimensional and involving not only income, but also food security, health, education, and gender equality. The project therefore lent itself to be evaluated using multiple indicators as well as using multidimensional poverty indices.

In this article we assess the impact of MVP on the MDGs indicators and on a poverty index, and then we compare the results. Many poverty indices have been proposed in the literature and in this paper we employ the global multidimensional poverty index of Alkire and Santos (2014) - global MPI for brevity - , which is based on the methodology for multidimensional poverty measurement outlined in Alkire and Foster (2011). The use of the global MPI has a number of advantages. First, it captures 'core' MDG indicators as it was built to align, to the extent it was possible, with indicators used to track the MDGs (Alkire and Jahan, 2018). Second, unlike ad-hoc indices, it produces results that are easily interpretable. The MPI was adopted by the UNDP for the measurement of global poverty in the Human Development Report series since 2010 (UNDP, 2010), and it is used for ranking countries and regions in national and international comparisons. Changes in the global MPI can be interpreted by benchmarking to changes in national trends. Finally, the MPI is theoretically grounded in the capability approach to poverty (Sen, 1992) and it is less affected by issues of arbitrariness and transparency that affect ad hoc indices.

established in 2010 with the goal of coordinating development efforts in the northern regions of Ghana. SADA was involved in the design and implementation of the MVP in Northern Ghana and the project came to be known as the SADA-Northern Ghana Millennium Village Project. SADA discontinued operations and ceased to exist in 2016.

We find that the MVP had a limited impact on the MDG indicators but a large impact on the global MPI. This result is worrying because the same program was rather ineffective when evaluated using multiple indicators, but turned out to be very effective when evaluated using the global MPI. Our sensitivity analysis led to a reconciliation of these contrasting results, and showed that the positive impact of MVP on the global MPI was mainly driven by improvements in just two MDG indicators: access to improved sanitation and primary school attendance. We conclude that the MVP was not a successful programme, and that the global MPI, in its current form, should be used with caution in the evaluation of development interventions and policies.

The paper is structured in the following way. The next section describes the SADA-MVP project, the evaluation design and the datasets. Section 3 presents the impact of the intervention on the MDG indicators. Section 4 illustrates the impact of MVP on the global MPI. Section 5 assesses the robustness of the results, and Section 6 concludes.

2 The SADA-Northern Ghana MVP evaluation

The SADA-MVP was implemented between May 2012 and December 2016 with funding from the UK Department for International Development (DFID). The project was implemented in a cluster of communities spread across the West Mamprusi and the Builsa districts of Northern Ghana. It provided a package of services in agriculture, health, education and infrastructure to 35 communities with an approximate population of 27,000 individuals.

More specifically, agricultural activities included: the promotion of farmers' based organizations and cooperatives; the provision of fertilizer, seeds and tractor services; training; agricultural loans; and construction of storage facilities. Activities in health included: the construction, rehabilitation and staffing of health clinics; the deployment of community health workers in home visits; the provision of basic drugs and preventative treatments (Vitamin A, deworming, iron, vaccines, anti-malarial drugs, and mosquito bed-nets); registration to the national health insurance scheme; and behavioural change campaigns. Activities in education included: the construction, rehabilitation and staffing of primary schools; teachers' training; construction of school toilets; scholarships for girls attending secondary school; social mobilization through parents and teachers' associations and school management committees; construction of teachers' accommodations; provision of basic school supplies; and establishment of IT learning centers. Infrastructural activities included the rehabilitation and construction of roads, boreholes and water points, home latrines, and the promotion of communication technology.

MVP had many critics as well as enthusiastic supporters. It was endorsed by the UN Secretary General, several prime ministers, philanthropists, academics, and celebrities. For 10 years it was the UN flagship anti-poverty programme. However, in the opinion of some researchers it was the quintessential expression of "philanthrocapitalism" (Wilson, 2015), a top-down and technocratic approach to development that disregards complexities and local communities. Munk (2013), mostly relying on anecdotal observations, ascribed the apparent failure of MVP to a simplistic approach to development that ignored the complexities of African poverty. White (2015) depicted MVP as a reproposition of old integrated development projects, which, like its predecessors, was bound to fail. A similar argument had been made earlier on by Clemens and Demombynes (2011), who more forcefully than others argued in favour of a rigorous impact evaluation of MVP.

Yet impact evaluations of millennium village projects have been few and of variable quality. Available evaluation studies assessed the impact of MVP on a limited number of MDG indicators, often using inadequate designs without baselines or control groups. Pronyk et al.(2012) found improvements on skilled birth attendance, bednet use, malaria incidence, access to sanitation and child mortality, but their study did not include a baseline and the impact on child mortality was challenged and retracted (Bump et al., 2012; Pronyk, 2012). Remans et al. (2011) found a reduction in stunting and improvements in indicators of food security, child care and infectious disease control, but they did not use a control group and compared beneficiaries before and after the intervention. Wanjala and Muradian (2013) found large impacts on maize yields, profits from maize production and consumption of own produced food in the village of Sauri in Kenya, but the study did not include a baseline and the sample was very small. Mitchell et al. (2018) examined data from project and control sites in 10 different countries, and found positive impacts on 30 out of 40 'MDGs-related' outcomes, particularly in the agriculture and health sectors, but the study did not include a baseline and many of the estimated outcomes were not MDG indicators. Michelson and Tully (2018) concluded that project effects in the Sauri village in Kenya were very small, but it relied on the use of land prices as an indicator of economic growth.

Our study is the only impact evaluation of MVP, which employs a baseline, a valid control group, and that assesses impact on all MDG indicators.[2] The evaluation design consists of a difference-in-difference (DiD) analysis of matched samples of villages (Masset et al., 2013). Before the project started, project villages were matched to an equal number of 'near' and 'far' control villages using village-level characteristics from the 2000 and 2010 population censuses.

Surveys were designed with the goal of tracking the largest possible number of MDG indicators. Baselines were conducted in 2012, and follow-up rounds were carried out every year from 2013 to 2016. The survey instruments included: a household questionnaire modelled to the Living Standard Measurement Surveys; an adult questionnaire modelled to the Demographic and Health Survey (DHS); anthropometric measurements of children under-5; blood specimens of children under-5 to assess the prevalence of malaria; and a community questionnaire. Table 1 shows the number of observations collected at each survey round by each survey tool. Attrition was very low and even at the end of the study did not exceed 5%. Attrition rates were very similar in the project and the comparison group and regression analysis showed that very few characteristics were correlated with attrition.

We estimated DiD effects at each survey round and over the full period. The first is

[2] The evaluation was registered with the Register for International Development Impact Evaluations (RIDIE) hosted by 3ie (Masset, 2015), and a pre-analysis plan was published before analysing the data (Masset, 2014).

Table 1: Observations by survey instrument and survey round

	2012	2013	2014	2015	2016
Households	2,172	2,230	2,191	2,177	2,185
Adult females (15-49)	2,837	-	3,241	-	2,837
Adult males (15-49)	1,628	-	1,835	-	1,671
Anthropometric measurements of children under-5	1,933	-	1,670	-	1,513
Blood samples of children under-5	805	-	1,121	-	968
Community surveys	103	103	103	103	103

the difference in the change in the outcomes between the baseline and each survey round in project and control villages. The second is the average of the impacts estimated at each round. In the regression analysis, we use cross-sectional or fixed-effects models depending on whether panel data were available. For some of the outcomes, for example, undernutrition, school attendance and mortality, panel data were not available because most children under-5 at the baseline were no longer in the sample at the endline. The cross-section regression model for the average treatment effect is:

$$y_{it} = \alpha + \gamma P_i + \delta P_i POST_i + \mu_t T_t + \beta X_i + \epsilon_{it} \tag{2.1}$$

where y_{it} is the outcome observed for the observation i at time $t = 0, \ldots, 4$ (where 0 is the baseline, and $t = 1, \ldots, 4$ are four following survey rounds). T_t is a vector of dummy variables for each follow-up period, P is a dummy variable equal to 1 if the observation is in the project group and equal to 0 if the observation is in the control group. $POST$ is a dummy variable equal to one for every observation collected after the baseline. The coefficient δ of the interaction of the project variable with the $POST$ variable is the average effect of the intervention. The X is a vector of baseline characteristics that improve the balance between the project and the control group.

The cross-section model for round-specific project effects is:

$$y_{it} = \alpha + \gamma P_i + \delta_t P_i T_t + \mu_t T_t + \beta X_i + \epsilon_{it} \tag{2.2}$$

where coefficients and variables have the same interpretation as before, except that there are now four different project effects, one for every survey round (δ_t).

With panel data we used a fixed-effects model to account for the impact of time-invariant household-level unobservable determinants. The average project effect and the round-specific effects were estimated using the following models respectively:

$$y_{it} = \alpha_i + \delta P_i POST_i + \mu_t T_t + \beta X_{it} + \epsilon_{it} \tag{2.3}$$

$$y_{it} = \alpha_i + \delta_t P_i T_t + \mu_t T_t + \beta X_{it} + \epsilon_{it} \tag{2.4}$$

The covariates (X) in this case are time-varying variables that are not affected by the project and include reported weather shocks (floods and droughts), and household demographic composition.

In order to reduce remaining imbalances between the project and control samples we further adjusted the estimation using the subclassification matching algorithm of Imbens and Rubin (2015). In the matching algorithm we included baseline values of household characteristics that were not affected by the intervention: household size; the age and education of the head of household; the size of cultivated land; an estimate of the value of all assets owned; whether the family received remittances from relatives; whether the household is farming and whether is farming millet, rice or groundnut; whether it had been affected by draught and flood shocks during the previous 5 years; the number of months it is normally food insecure; whether it has access to bank loans; whether the house has a metal roof; and the distance to drinking water. The same covariates were also used for estimating project effects using a simpler procedure ("regression adjustment"), whereby DiD effects are estimated controlling for baseline characteristics, whose results are shown in Tables 8, 9 and 10 in the Appendix.

The validity of DiD analysis rests on the similarity of the trends in the outcomes in the project and control groups. In order to assess the validity of this "parallel trends" assumption, we collected retrospective data on school attendance, livestock holdings and cultivated land. We tested for differences in the trends up to two years before the baseline and we found none. Lastly we proved the validity of our estimation strategy with a placebo test. We estimated the impact of a "pseudo-intervention" that did not occur. If our estimation strategy is correct, we should not observe impacts of interventions that did not take place. We assessed the impact of an hypothetical intervention in the Builsa 'far' communities using the West Mamprusi 'far' communities as the control group. This test is very conservative because the two districts are different in many ways and expected to diverge over time. We found very small impacts, which were statistically significant only in two cases: the proportion of children sleeping under bednets, and household access to sanitation.[3]

A final concern of our study design is the multiple hypotheses testing problem. When we conduct many statistical tests at the same time, some null hypotheses will be rejected by chance alone. For example, with 29 hypotheses about impacts on MDG indicators, and with a statistical significance threshold of 10%, there is a 95% probability of finding at least one statistical significant effect even if the intervention has no impact (the probability is $1-(1-\alpha)^N$, where α is the level of statistical significance and N is the number of hypotheses). Approaches to addressing the multiple testing problem rely on more stringent requirements for declaring statistical significance. For example, in one popular approach (the *Bonferroni* method), the critical value to establish statistical significance is set to α/N (with 10% significance and 29 hypotheses, the critical value is 0.0034).

A less conservative approach is a modification of the False Discovery Rate (FDR) proposed by Anderson (2008). This approach follows a two-step procedure. In the first step, an algorithm orders the p-values of each test in ascending order indexing them by $i = 1, \ldots, N$ and rejects all the null hypotheses whose p-value is less than $\frac{i}{N}\alpha$. In second step the pro-

[3]A more detailed description of the test for the presence of parallel trends and of the placebo test can be found in (Masset et al., 2020).

cedure is implemented again using in the denominator, rather than N, the original number of hypothesis minus the number of hypothesis rejected in the first stage. In the tables with the estimation results, we mark with a star the coefficients whose p-values are below the statistical significance defined by the Anderson FDR with $\alpha = 10\%$. Coefficients marked by a star can be interpreted as the remaining statistically significant results after removing false discoveries. Strictly speaking, the stars do not correspond to standard statistical tests and can be defined as 'interesting' results that deserve our attention (Efron and Hastie, 2016).

3 The impact of the SADA-MVP on the MDG indicators

We first assess the impact of MVP on the MDG indicators.[4] The MDGs are tracked by 60 indicators and our study assessed impact on 29. Twenty-one of the 60 indicators are measured at the national or international level and cannot be calculated using household level data.[5] Ten indicators could not be estimated with our data because the required information was not collected or because the samples were too small to estimate averages and performing statistical tests.[6] We calculated values of these 29 indicators following the UN guidelines for

[4]A full evaluation of the SADA-Nothern Ghana MVP including an analysis of spill-over effects, impacts on interventions in neighbouring areas, cost-effectiveness analysis, and robustness checks can be found in Masset et al. (2020)

[5]The MDG indicators that cannot be calculated at household level are: 3.3 Proportion of seats held by women in national parliament, 7.1 Proportion of land area covered by forest, 7.2 CO2 emissions, total, per capita and per $1 GDP (PPP), 7.3 Consumption of ozone-depleting substances, 7.4 Proportion of fish stocks within safe biological limits, 7.5 Proportion of total water resources used , 7.6 Proportion of terrestrial and marine areas protected, 7.7 Proportion of species threatened with extinction, and 7.10 Proportion of urban population living in slums. Similarly, 12 other outcomes (8.1 through 8.12) relate to official development assistance, market access, and debt sustainability and can only be calculated at national or international level.

[6]The household-level indicators that could not be measured with the available data are: 5.1 Maternal mortality rate, 5.6 Unmet need for family planning, 6.1 HIV prevalence among population aged 15-24 years, 6.2 Condom use at last high-risk sex, 6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10-14 years, 6.5 Proportion of population with advanced HIV infection with access to antiretroviral drugs, 6.9 Incidence, prevalence and death rates associated with tuberculosis, 6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course, 8.13 Proportion of population with access to affordable essential drugs on a sustainable basis, and 8.16 Internet users per 100 inhabitants.

monitoring the MDGs.[7] A full list of the MDGs indicators including a description of how they were measured and of how they should be interpreted is reported in Table 3.

The estimated average impact of the intervention on each MDG indicator and the year-specific impacts are reported in Table 2. The trajectories of the indicators over time in project and control areas are depicted in Figures 1 and 2. In all charts, MV stands for millennium villages and CV stands for control villages.

[7]Available at http://mdgs.un.org/unsd/mdg/host.aspx?Content=indicators/officiallist.htm

Table 2: Impact of MVP on the MDG indicators

MDG	DD Impact 2013	DD Impact 2014	DD Impact 2015	DD Impact 2016	DD Average impact
Proportion of population below $1.25 (PPP) per day	-11.70*	-9.84*	-3.94	-9.04*	-8.66*
	(3.67)	(3.96)	(3.77)	(3.70)	(2.73)
Proportion of population below the national poverty line	-1.39	-1.46	2.07	5.76	1.16
	(3.72)	(3.99)	(4.31)	(4.58)	(3.41)
Poverty gap ratio	-0.15	-3.98	4.51	2.63	0.71
	(2.79)	(2.98)	(2.61)	(3.30)	(2.32)
Consumption share of poorest quintile	1.17	1.49	0.99	-0.01	0.87
	(1.26)	(1.37)	(1.17)	(1.16)	(0.86)
Employment to population ratio	2.15	5.22	4.31	0.80	3.06
	(2.84)	(2.93)	(2.49)	(3.15)	(2.41)
Proportion of employed people living below $1 (PPP) per day	-13.59*	-13.63*	-6.60	-10.01*	-11.04*
	(4.22)	(4.18)	(4.09)	(3.74)	(3.03)
Proportion of own account and contributing family workers in total employment	3.89	3.90	4.28	4.04	4.02
	(1.92)	(1.96)	(1.97)	(2.06)	(1.93)
Underweight prevalence (children under-5]		1.03		-2.14	-0.52
		(2.94)		(2.75)	(2.28)
Proportion of population below the food poverty line	-0.78	-5.55	10.63*	1.67	1.42
	(4.54)	(5.11)	(4.67)	(4.85)	(3.84)
Net attendance ratio in primary education	9.56*	4.35	3.54	13.48*	7.69*
	(3.53)	(3.80)	(3.59)	(3.80)	(3.16)
Completion rate in primary education	0.90	-1.43	-1.40	-4.12	-1.62
	(4.38)	(4.06)	(4.25)	(3.98)	(3.79)
Young adults (15-24] literacy rate		-6.19		-0.19	-3.36
		(3.91)		(3.88)	(3.33)
Ratio of girls to boys in primary education	-28.71*	-8.75	-10.34	-25.93	-18.70
	(10.88)	(10.41)	(12.77)	(11.87)	(9.68)
Share of women employed in the non-agricultural sector	-10.97	0.96	-6.92	-14.54	-8.06
	(17.52)	(19.03)	(15.91)	(16.77)	(13.40)
Under-5 mortality rate		-2.31		0.41	0.41
		(2.41)		(2.07)	(2.07)
Infant mortality rate		-1.09		2.02	2.02
		(2.33)		(1.89)	(1.89)
Measles immunisation rate (children under-2]		-6.45		-3.10	-4.95
		(4.59)		(5.12)	(3.70)
Proportion of births attended by skilled health personnel		16.57*		39.08*	27.00*
		(4.83)		(5.90)	(4.50)
Contraceptive prevalence rate		5.36*		11.26*	8.22*
		(2.33)		(2.68)	(2.13)
Adolescent birth rate				-8.67	-8.67
				(8.31)	(8.31)
Ante-natal care coverage		-7.43		2.36	-2.94
		(4.89)		(3.83)	(4.06)
Proportion of young adults (15-24] with correct HIV knowledge		0.56		2.41	1.47
		(2.65)		(2.09)	(2.05)
Malaria prevalence (children under-5]		-4.50		-4.47	-5.53
		(4.65)		(4.74)	(4.28)
Proportion of children under-5 sleeping under insecticide treated bed-nets		42.88*		34.60*	39.24*
		(6.13)		(6.73)	(5.85)
Proportion of children under-5 with fever treated with anti-malarial		12.20		24.25*	16.65
		(9.72)		(10.54)	(8.52)
Proportion of the population using an improved drinking water source		-6.84		5.92	-0.61
		(4.29)		(4.24)	(3.83)
Proportion of the population using an improved sanitation facility		1.22		60.69*	30.33*
		(2.16)		(5.46)	(3.16)
Fixed telephone subscriptions rate		0.01		0.07	0.04
		(0.16)		(0.11)	(0.12)
Mobile telephone usage rate		-5.40		-9.96	-7.60
		(6.08)		(5.27)	(5.23)

Note: Coefficients are difference-in-difference estimates expressed in per cent terms (ie the coefficient are multiplied by 100). Standard errors reported in parentheses are calculated using 500 bootstrap replications. A star * represents a statistically significant coefficient at 10% after applying the False Discovery Ratio sharpened q-values following Anderson (2008). Infant and under-5 mortality rates were calculated using the DHS synthetic cohort probability method using the SYNCMRATES stata package. Standard practice is calculating mortality rates over an interval of five years before the survey. As a result, the endline and the average changes in mortality rates are the same.

All MDG indicators are binary, and the impact estimates can be interpreted as percentage points differences between the project and the control group. The project had a beneficial impact on seven MDG indicators: income poverty, employment rate of people below the income poverty line, net attendance rate in primary school, the proportion of births attended by skilled professionals, the proportion of women using contraceptive methods, the proportion of children under-5 sleeping under bed nets, and the proportion of the population with access to improved sanitation. The impacts on the use of mosquito bed-nets and on access to sanitation were large, while other impacts were small (the difference was below 10 percentage points). There were no negative impacts, with the possible exception of a reduction of the gender parity ratio in primary education.

The project produced, at best, mixed results. Even excluding some indicators that were not explicitly targeted by the intervention, such as adult literacy rates and the number of landline connections, the project only improved about 25% of the MDG indicators (seven out of 27). Agricultural interventions brought about a reduction in income poverty and an increase in employment, but expenditure poverty remained unchanged. Interestingly, the project reduced income poverty (based on a $1.25 purchasing power parity poverty line), but did not reduce expenditure poverty (based on per-adult equivalent consumption).[8] Education activities produced an increase in primary school attendance, but the project failed to promote school retention as measured by completion rate in primary school. Health interventions made considerable progress in fighting malaria and in providing perinatal and post-natal care. However, progress in intermediate outcomes did not affect final outcomes: child mortality, undernourishment, and malaria incidence. Finally, infrastructural interventions gave households access to sanitation facilities, but there was no improvement in access to drinking water, or in mobile and internet technology. The project did not perform better

[8]Further analysis of the data shows that households increased assets holdings during the project suggesting that income gains were invested rather than spent. This result is consistent with the consumer behaviour predicted by the permanent income hypothesis. Consumers interpreted income changes brought about by the intervention as temporary rather than permanent, and did not therefore adjust their expenditure levels, a phenomenon also observed in the evaluation of other poverty eradication programmes (Ravallion and Chen, 2005).

in terms of equity goals. MVP had a focus on extreme poverty and on gender but did not affect the distribution of expenditure nor it improved gender empowerment.

As for the reasons for the lack of impact, these are difficult to identify. The project did not count with a clear theory of change mapping how activities would affect outputs and final outcomes. The design of such a theory of change proved to be an incredibly difficult task, given the number of activities implemented and their interrelations. We therefore opted for evaluating the intervention as a "package", rather than attempting a separate evaluation of its single activities. A quick look at the impacts on the MDG indicators shows that the project was relatively effective in improving outputs, such as, for example, the use of bednets, access to sanitation, and births attended by skilled professionals. Final outcomes, however, such as child mortality and undernutrition, remained unaffected, suggesting the presence of constraints to the transformation of outputs into outcomes. It is possible that the design and the implementation of a large number of interrelated activities led to the disregard of the contextual characteristics that were more conducive to the success of the promoted activities. There is a sizeable literature on the failure of other integrated rural development programmes (Masset, 2018), whose lessons might be applicable to MVP. A summary of these lessons would include, in increasing order of importance: the selection of intervention areas with limited growth potential, the use of inappropriate top-down approaches to project design, and the difficulty to manage and implement multi-sector interventions of exceeding complexity.

Figure 1: Trends of MDG indicators in project and control areas

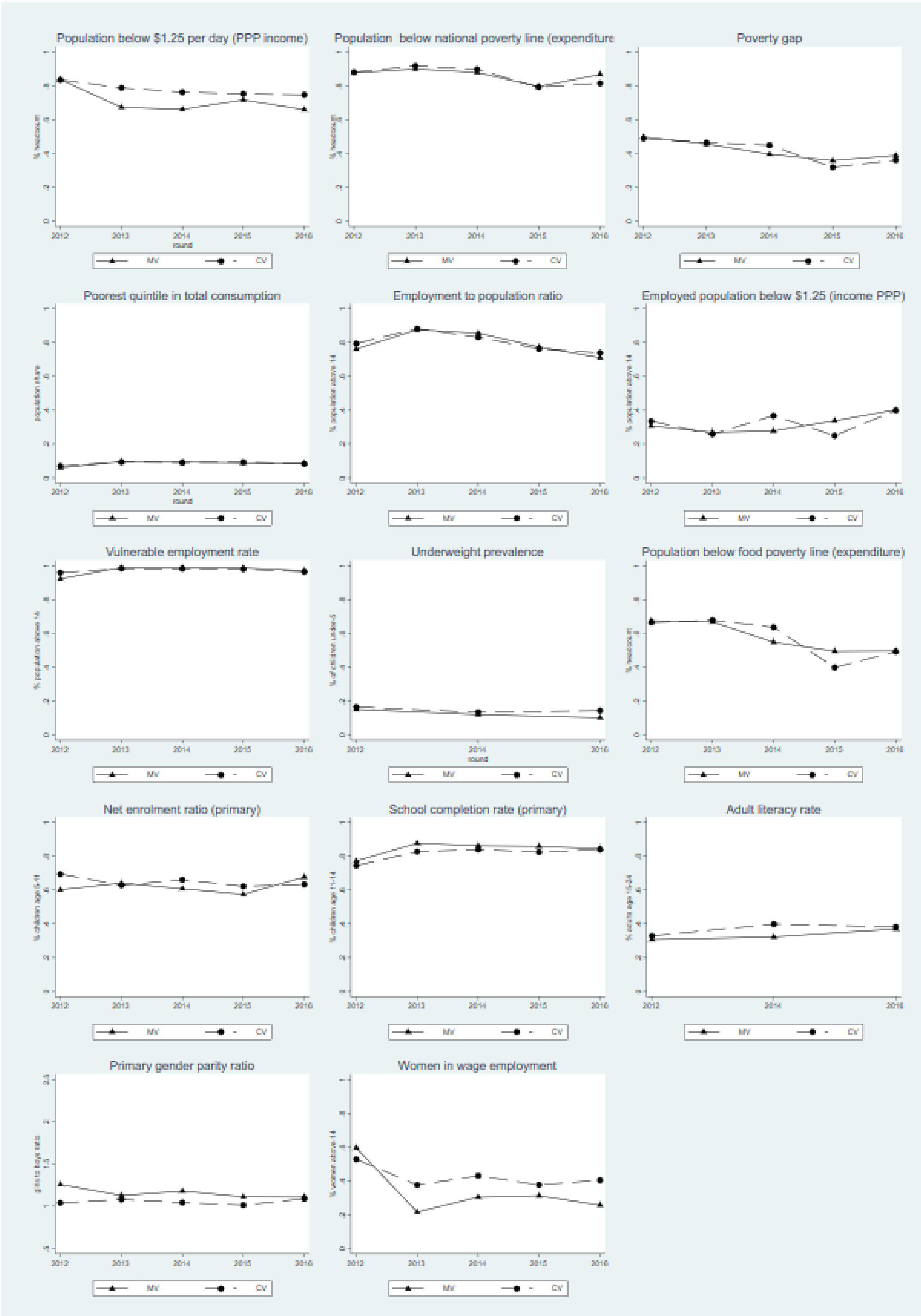


Figure 2: Trends of MDG indicators in project and control areas

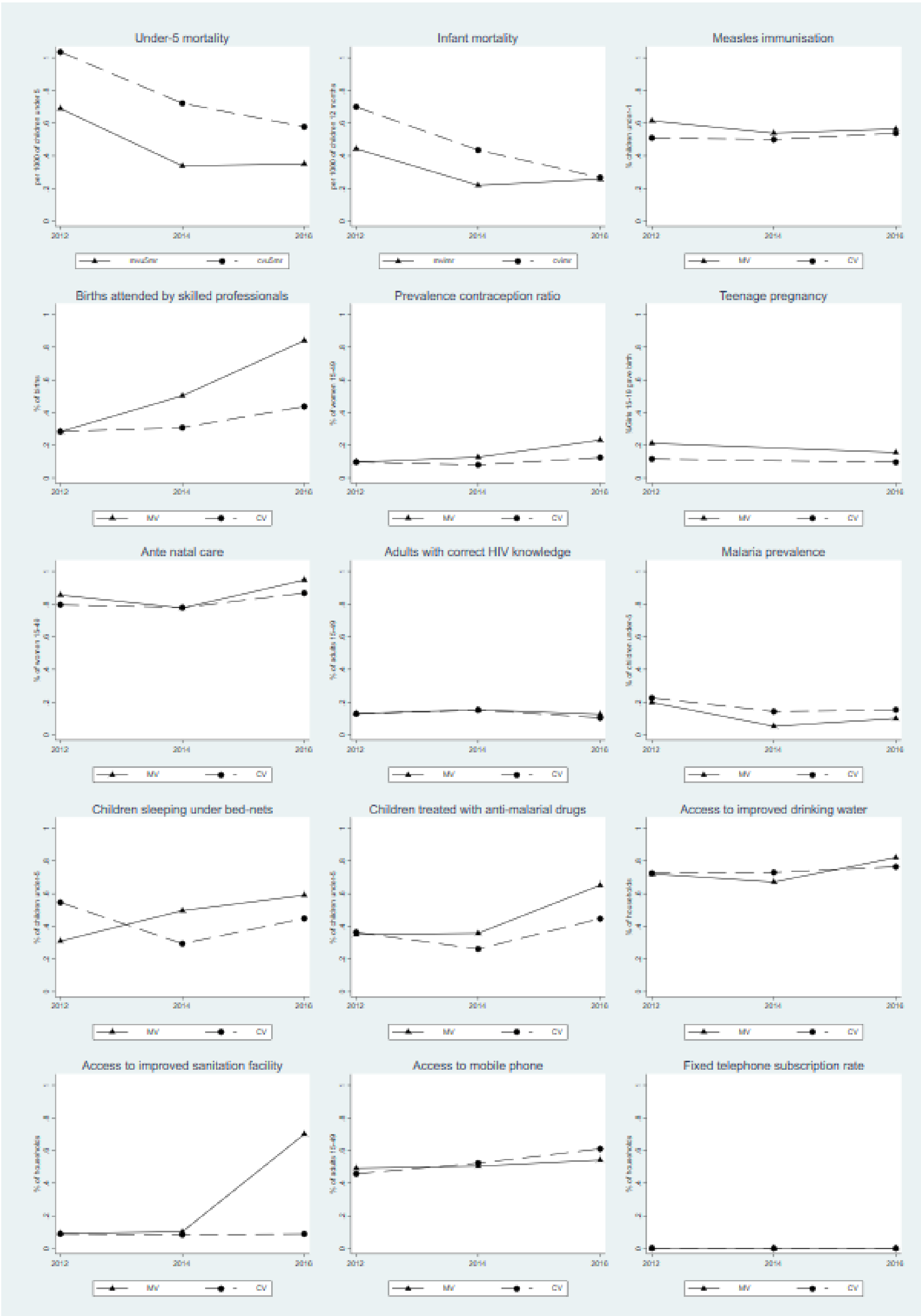


Table 3: The MDG indicators

MDG indicators	Obs.	Panel	Description
1.1 Proportion of population below $1 (PPP) per day	5	Yes	The proportion of the population whose per-capita income falls below the international poverty line of $1.25 a day at purchasing power parity, thus adjusting for cost of living in Ghana.
1.2 Proportion of population below the national poverty line	5	Yes	The proportion population whose per-adult equivalent expenditure falls below the official Ghanaian national poverty line allowing the purchase of a minimum basket of food and non-food items.
1.3 Poverty gap ratio	5	Yes	The mean shortfall in the population from the national poverty line. It measures the depth of poverty by calculating how far the poor are from the poverty line.
1.4 Share of poorest quintile in national consumption	5	Yes	The share of total expenditure in the study area that goes to the poorest 20% of the population. This is a measure of expenditure inequality in the population.
1.5 Employment to population ratio	5	No	The percentage of individuals older than 15 who did any work, paid or unpaid, over the previous year not including domestic work.
1.6 Proportion of employed people living below $1 (PPP) per day	5	No	The percentage of the employed (as defined above) who are poor (as defined by the income $1.25 poverty line). This indicator measures the availability of decent work.
1.7 Proportion of own account and contributing family workers in total employment	5	No	The proportion of the employed population (as defined above), engaged in farming, animal husbandry, fishery or any other self-employment without being remunerated.
1.8 Percentage of underweight children under-5	3	No	The percentage of children aged 0-59 months, whose weight is below the WHO benchmark. It is a composite indicator of acute (wasting) and chronic (stunting) malnutrition.
1.9 Proportion of population below a minimum level of dietary energy consumption	5	Yes	The proportion of individuals whose per-adult equivalent expenditure falls below the Ghanaian official food poverty line which allows the purchase of a minimum basket of food items.
2.1 Net enrolment ratio in primary education	5	No	The proportion of children of official primary school age (6-11) that are reported having attended primary school at any time during the previous year (also known as Net Attendance Ratio).
2.2 Proportion of pupils starting grade 1 who reach last grade of primary	5	No	The proportion of children age 11-14 who completed primary school among those who ever attended primary.
2.3 Literacy rate of 15-24 year olds, women and men	3	No	The proportion of male and female adults age 15-24 who were able to read correctly two English sentences ("The child is playing with the ball", and "Farming is hard work") and to solve basic arithmetic (9+4 and 4x5).
3.1 Ratio of girls to boys in primary education	5	No	The ratio of the net attendance rate in primary school of boys and girls age 6-11. A ratio below one implies fewer girls are attending primary than boys.
3.2 Share of women in wage employment in the non-agricultural sector	5	No	The proportion of women above 15 in overall employment in the non-agricultural sector. The indicator measures to what extent women have equal access to jobs outside agriculture.
4.1 Under 5 mortality rate	2	No	The probability of a child dying before age 5, calculated per thousand of the population over the 5 years preceding the interview using the DHS synthetic cohort probability method.
4.2 Infant mortality rate	2	No	The probability of a child dying before 12 months of age, calculated per thousand of the population over the 5 years preceding the interview using the DHS synthetic cohort probability method.
4.3 Proportion of 1-year-old children immunised against measles	3	No	The proportion of children aged 0 or 1 whose vaccination card reports a measles vaccination or whose mother recall the child being given an injection in the upper arm to prevent measles.
5.2 Proportion of births attended by skilled health personnel	3	No	The proportion of deliveries assisted either by doctor, clinical officer, nurse, or midwife for all children of age 0-2 at the time of the interview.
5.3 Contraceptive prevalence rate	3	No	The proportion of women aged 15-49 using any contraceptive method at the time of the interview (sterilisation, pill, IUD, injections, implants, condoms, rhythm, abstinence, or withdrawal).
5.4 Adolescent birth rate	2	No	The proportion of women aged 15-19 that gave birth during the previous 5 years.
5.5 Antenatal care coverage	3	No	The percentage of women aged 15-49 who received at least one antenatal visit (by a doctor, clinical officer, nurse, midwife, or CHW) for children who were aged 0-2 years at the time of the interview.
6.3 Proportion of population aged 15-24 with comprehensive correct knowledge about HIV	3	No	The proportion of population aged 15-49 that answered correctly 5 (yes/no) questions about obvious causes of HIV infection transmission.
6.6 Malaria prevalence among children under-5	3	No	The proportion of children under-5 with malaria based on microscopic analysis of parasite count in the blood
6.7 Proportion of children under-5 sleeping under insecticide treated bed nets	3	No	The proportion of children aged 0-59 months who slept under an insecticide-treated mosquito net the night before the interview.
6.8 Proportion of children under 5 with fever who are treated with anti-malarial drugs	3	No	The proportion of children aged 0-59 months with fever in the last 2 weeks who received anti-malarial drugs.
7.8 Proportion of the population using an improved drinking water source	3	Yes	The percentage of households whose main source of drinking water is one of the following: piped into welling, yard or plot; public tap; tube well and borehole; protected dug well; protected spring; bottles; sachet water.
7.9 Proportion of the population using an improved sanitation facility	3	Yes	The percentage of households that normally uses one of the following sanitation facilities: flush to piped sewer system; flush to septic tank; flush to pit (latrine); ventilated improved pit latrine; pit latrine with slab.
8.14 Fixed telephone subscriptions for 100 inhabitants	3	Yes	Percentage of households reporting having a landline in the home.
8.15 Mobile cellular subscriptions for 100 inhabitants	3	Yes	Percentage of adults aged 15-49 reporting personal use of a mobile phone during some or all the year before the interview.

Note: The table describes the MDGs indicators used in the study and indicates whether the indicator was observed for the same household/individual over time (panel column), in which case a fixed effect model was used to estimate project effects. The table also indicates the number of survey rounds available for each indicator (obs. column). Some outcomes were observed every year, like poverty or school attendance, while others were only observed at the baseline, the midterm and the endline, others still were only observed at baseline and endline because the survey questions refer to 4 or 5 years before the survey (child mortality and adolescent birth rates).

4 Impact of the SADA-MVP on the global MPI

The analysis of multiple indicators does not provide an overall assessment of the programme when results are mixed. MVP produced some positive effects in some areas but was ineffective in others, and the size of the effects varied across MDG indicators. One approach to assessing the overall impact of the intervention consists of counting the proportion of positive effects (see for example Mitchell et al.(2018)). This ”vote counting” approach, however, can be misleading because it focuses the attention on the statistical significance of the estimated effects rather than on their practical significance. The absence of statistical significance is not necessarily evidence of absence of a project effect and could be the consequence of low statistical power associated to a small sample size. On the other hand, a highly 'statistically significant' result can be very small and practically irrelevant. A second issue with this ”vote counting” approach is that some MDG indicators represent the same construct and tend to vary together (for example the proportion of population below a per capita income of $1.25 PPP and the proportion of the employed population below a per capita income of $1.25 PPP), thus creating a risk of double counting. Finally, counting the statistically significant effects implicitly assumes that all indicators have the save relevance, and it is equivalent to attributing each outcome the same importance. For all these reasons, counting the proportion of statistically significant results is not a valid approach at synthesising results of multiple indicators.

Conversely, indices synthesise multiple indicators in a single metric and allow straightforward comparisons between the project and the control group, which are not open to different interpretations. In addition, indices eliminate the multiple hypotheses testing problem by reducing all hypotheses to just one. The use of indices of multiple indicators has indeed become common in the evaluation literature (see for example Loschman et al. (2015)).

In our study, we could not build an ad hoc index of the 29 available MDG indicators for several reasons. First, the MDG indicators are calculated over different segments of the population (households, adult women, mothers, children under-5, adolescent girls, young adults, children under-2, women in reproductive age, children under-1, and school age children), in such a way that they cannot be aggregated across households. Second, some MDG indicators are population-level rates (the poverty gap, the share of consumption of poorest quintile, and the gender parity ratio), which cannot be calculated at the household level. Third, an ad hoc index would face decisions regarding the weights of each indicator, and the gain in the interpretation of results would come at the cost of a lack of transparency. Finally, the impact of the intervention on an ad hoc index is difficult to interpret. How can we say whether a change in the index is small, medium or large, without reference to changes in the same index observed by other studies?

Hence we opted for using a popular multidimensional poverty index: the global MPI. The global MPI was co-designed and launched in 2010 by the UNDP, the Human Development Report Office, and the Oxford Poverty and Human Development Initiative (OPHI), and by design it was aligned, to the extent it was possible, to the indicators used to track the MDGs (Alkire and Jahan, 2018). Although the global MPI is not an index of all the MDG indicators, the overlap is substantial. Of its 10 constitutive elements, six are also MDG indicators and measured in a nearly identical way (primary completion rate, primary attendance rate, proportion of underweight children, child mortality, access to improved sanitation, and access to drinking water), while the remaining four (access to electricity, use of dug floor, type of cooking fuel, and assets ownership) are set to capture living standards in the absence of poverty data.

A great advantage in the use of the global MPI is the greater interpretability of the results. The global MPI was adopted by the UNDP in 2010 for the measurement of global poverty, and data for its calculation are readily available from the Demographic and Health Surveys. The availability of historical series of the global MPI for most countries, and the easiness of calculation, allows researchers to interpret the impact of interventions by benchmarking. We are able to compare the impact of the MVP on the global MPI to changes that have occurred over time in the country and in other contexts. Similarly, the global MPI can be used to assess impact of different interventions in such a way that most successful poverty eradication interventions could be identified, an opportunity precluded by the use of ad hoc indices.

The global MPI is constructed in the following way. It considers deprivations along three dimensions: health, education, and living standards. These three dimensions are given equal importance (1/3) and indicators are calculated for each dimension. In particular, a household is deprived if: no household member has completed 5 years of schooling (with weight of 1/6), any school-age child is not attending school in years 1 to 8 (with weight of 1/6), any child has died in the family (with weight of 1/6), any child for which there is information is underweight (with weight of 1/6), the household has no electricity (1/18), the household does not have access to an improved sanitation facility (with weight of 1/18), the household does not have access to improved drinking water (with weight of 1/18), the household has dirt, sand or dung floor (with weight of 1/18), the household cooks with dung, wood or carbon (with weight 1/18), the household does not own more than one radio, TV, telephone, bike, motorbike or refrigerator, and does not own a car or truck (with weight of 1/18). The weighted sum of the deprivation indicators produces a deprivation score with values between 0 and 1 for each household.

We were able to calculate the index in the same way with just two minor exceptions. First, our malnourishment deprivation index is based on child undernutrition only, because our surveys did not measure the nutritional status of mothers, through the body-mass index. Second, we restricted the time over which calculating child mortality to 5 years before the survey in order to be able to measure more accurately changes produced by the intervention.[9] Since not all households have children under-5 or children in school age, some deprivation indicators are censored. We deal with censored observations following the same procedure adopted for the global MPI (Alkire and Santos, 2014): we consider not deprived those household for which no information is available.

[9]The latest version of the global MPI of 2018 has also adopted the same 5-year convention for the calculation of mortality deprivation (OPHDI, 2018).

Following Alkire and Santos (2010), we used the household deprivation score to calculate the following poverty indices :

- The multidimensional poverty headcount ratio, also called the **incidence** of poverty (H). This is the fraction of the population with a deprivation score equal or larger than one third.
- The average deprivation score among the poor, or the **intensity** of poverty (A).
- The average deprivation score of the poor across the whole population, also called the adjusted multidimensional poverty index (MPI). This is the **global MPI**, which is the average deprivation score after setting to zero the deprivation scores of households that are not multidimensionally poor according to the index H above. It can be shown that the global MPI is the product of the other two indices: the incidence and intensity of poverty (MPI=H*A)

A distinctive feature of the global MPI is that it reflects overlapping deprivations suffered by households at the same time. This feature is a consequence of the dual cut-off method used in its construction (Alkire and Foster, 2011). First, each household is classified as poor or not in each deprivation. Second, each household is classified as multidimensionally poor if it is deprived in at least one third of all deprivations. Because of the dual cut-off the index increases when deprivations become concentrated in the same households. The index is aligned with a system of social preferences that gives more weight to overlapping deprivations (Aaberge and Brandolini, 2015), which is closer to how most people understand poverty.

Our analysis shows that the impact of MVP on the global MPI was positive. The charts in Figure 3 show the impact of MVP on the global MPI, on multidimensional poverty incidence and intensity. The indices were nearly identical at the baseline in the project and control groups and a t-test showed that the differences were not statistically significant. After the intervention, multidimensional poverty decreased in the control group, but it decreased at a faster rate in the project group.

Figure 3: Impact of MVP on the global MPI

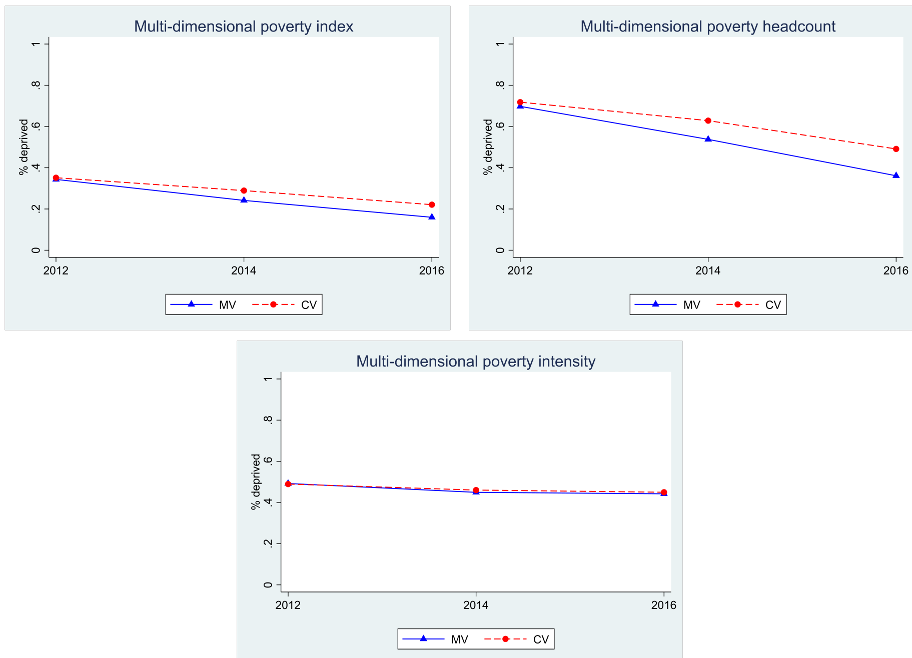


Table 4 presents the DiD impact estimates. MVP produced a statistically significant reduction in the global MPI and in poverty incidence. The impact was larger at the endline than at the midterm, pointing to a continuous impact of the intervention over time. Finally, the project decreased poverty intensity only marginally, and the effect was not statistically significant. The different impact of MVP on incidence and intensity of poverty is interesting from a policy perspective. Recall that the global MPI is the product of poverty incidence and of poverty intensity. If the global MPI is reduced via a large reduction in poverty incidence and a small reduction in poverty intensity, it means that the improvement occurred by improving the conditions of the poor with lower intensity of poverty. Conversely, changes relatively more favourable to the poorest of the poor would improve the global MPI by reducing relatively more poverty intensity than poverty incidence.

Was the impact of MVP on the global MPI large? We assessed the relevance of the impact by benchmarking the results. We compared poverty incidence in project and control

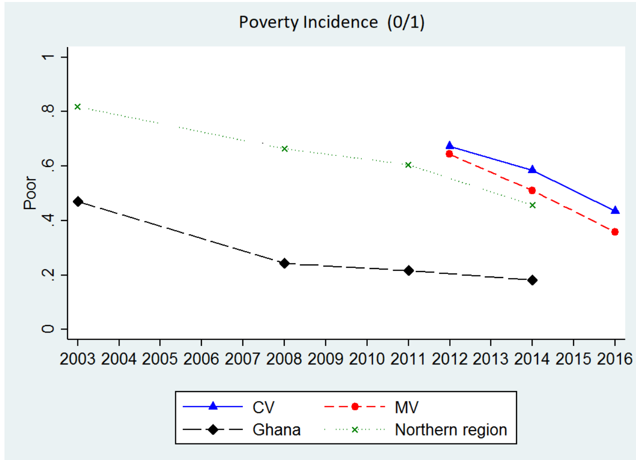
Table 4: Impact of MVP on multidimensional poverty

	Baseline in CV areas	Baseline difference in MV	DD impact 2014	DD impact 2016	Average DD impact
Multidimensional poverty index	35.16	-0.81	-3.64*	-5.48**	-4.56**
		(3.06)	(2.01)	(2.30)	(1.95)
Multidimensional poverty incidence	71.82	-2.04	-6.42*	-11.37**	-8.90**
		(4.85)	(3.70)	(4.66)	(3.67)
Multidimensional poverty intensity	48.96	0.24	-1.74	-0.96	-1.50
		(1.32)	(1.11)	(1.33)	(0.95)

Note: Coefficients are difference-in-difference estimates using a cross-sectional model estimated using sub-classification on a trimmed sample. Standard errors calculated using 500 bootstrap replications. Standard errors reported in parentheses are clustered at the village level. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%

areas to trends in Ghana and in the Northern region (see Figure 4). At baseline, poverty incidence in the study area was four times the level in Ghana, and it was larger than in the Northern Region. After the intervention, progress occurred in both project and control areas, but at different rates. After four years, the gap of the project area with the rest of Ghana decreased by 50%, while the gap reduction in the control area was only 30%. By the end of the project, the MVP area caught up with the Northern Region, while the control area did not. At current trends, the control areas will catch up with the rest of the country in 2024 while the project areas would catch up in 2022, implying the MVP produced a two year acceleration in current trends.

Figure 4: Multidimensional poverty in Ghana and in the study area

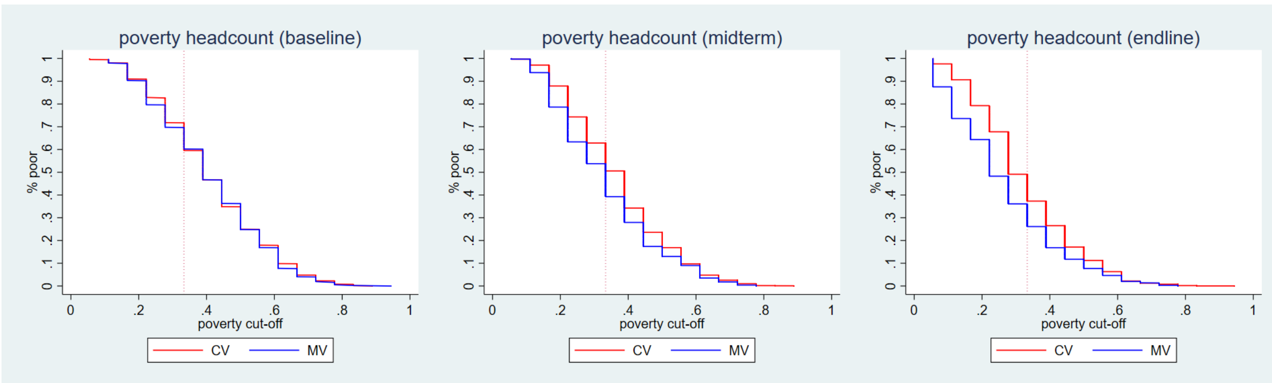


5 Sensitivity analysis

In this section we assess the robustness of the MPI impact estimates with the goal of reconciling the contrasting results obtained in the previous sections. In particular, a) we assess the sensitivity of the global MPI to the choice of the poverty cut-off, b) we analyse to what extent the changes observed in the global MPI are driven by changes in one or two of its components, and c) we measure to what extent they reflect changes in the distribution of deprivations in the population.

We assessed the sensitivity to the poverty cut-off with simple stochastic dominance analysis. The global MPI employs a 1/3 cut-off, meaning that a household (and all its members) are poor if their deprivation score is equal or larger than a third. It would be problematic if we obtained a different result using a different cut-off of, for example, 1/2 or 1/5. The charts in Figure 5 plot poverty incidence for all possible poverty lines for the three survey rounds separately. When the poverty line is 0, everybody is poor. Poverty decreases as we increase the poverty cut-off, and when the cut-off is 1 (a household has to be deprived in all dimensions to be classified as poor) very few households are poor. At baseline, the poverty distributions in the project and control areas overlap and poverty incidence is nearly identical for all possible poverty lines. At the midterm and endline poverty is unequivocally lower in project areas. At no poverty cut-off the lines are crossing, and it cannot be argued that impact is particularly large at the 1/3 cut-off. We conclude that the impact of MVP on multidimensional poverty was independent of the poverty cut-off.

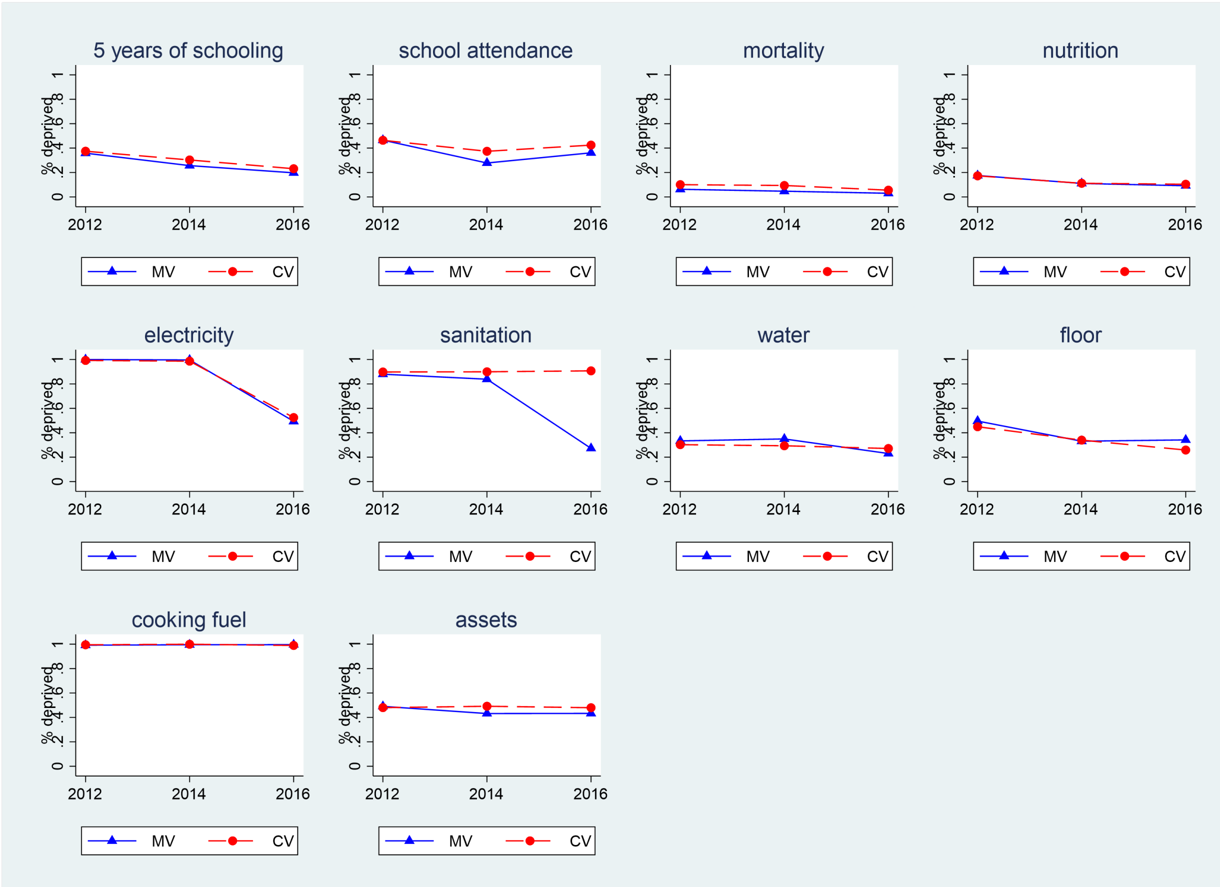
Figure 5: Stochastic dominance analysis



The global MPI is an index of 10 deprivations and it is useful to see in what dimensions the improvement occurred. We broke down the index by deprivation and analysed the impacts separately. The charts in Figure 6 show the percentage of the deprived population for each deprivation in project and control villages over time. The figure indicates a clear impact on sanitation, but no other large effects are visible.

We tested the impact of MVP on each deprivation using DiD analysis (see the results in Table 5). Deprivations were fairly similar at the baseline in the project and control group with only one indicator (child mortality) showing a statistically significant difference. The project reduced all deprivation indices, with the exception of child mortality and the use of cooking fuel. The impacts however were small (less than 10 percentage points), with the exception of a large impact on sanitation, and they were statistically significant only in the case of school attendance and sanitation.

Figure 6: MPI deprivation trends in project and control areas



The third column of Table 5 shows the contribution of each deprivation to the global MPI at the baseline. Contributions are obtained multiplying each censored deprivation (setting to zero the deprivation index for households that are not multidimensionally poor), by the weight assigned to each deprivation, and dividing by the global MPI. The percentage contributions tell us what are the main drivers of overall deprivation. At the baseline, nearly 40% of total deprivation was caused by failures in education. Another 20% was driven by failures in sanitation and electricity. One interpretation of the large impact of MVP on the global MPI is that it had a large impact on two deprivations (school attendance and sanitation), which together accounted for more than 30% of total deprivation.

We further investigated the sensitivity of the results by estimating impacts without each indicator in turn. We estimated poverty after leaving out one of the index components at

Table 5: Impact of MVP on deprivation indices

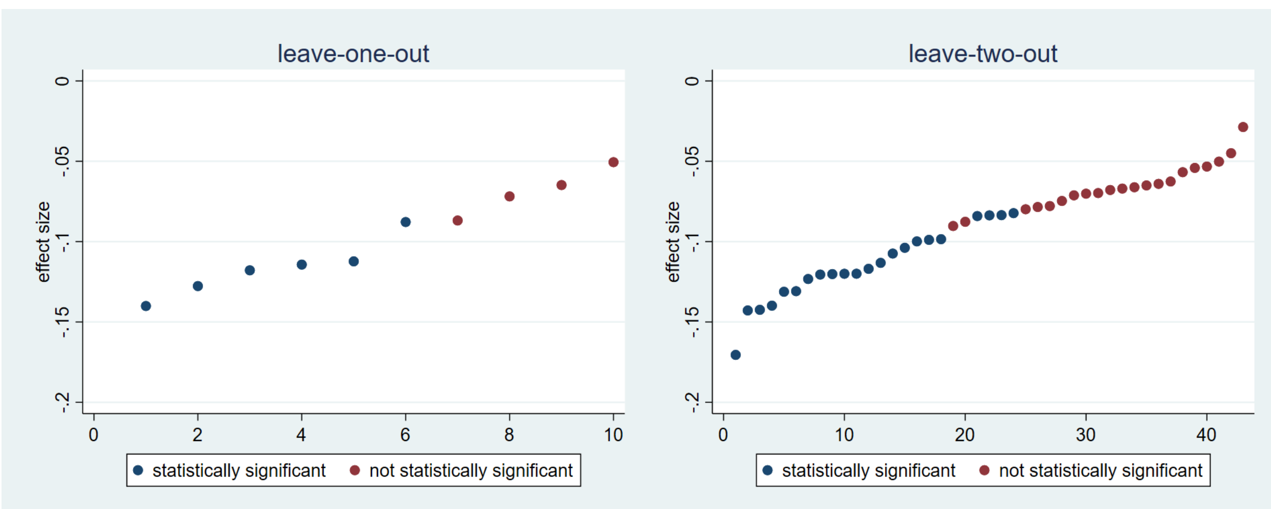
	Baseline in CV areas	Baseline difference in MV	contribution to index	DD impact 2014	DD impact 2016	Average DD impact
Years of schooling	37.46	-1.60	17.03	-2.82	-1.78	-2.30
		(5.66)		(3.28)	(3.16)	(3.01)
School attendance	46.39	0.11	22.22	-9.13**	-6.49	-7.83**
		(4.58)		(3.61)	(4.10)	(3.40)
Child mortality	10.07	-3.80**	2.99	-1.18	1.51	0.15
		(1.63)		(2.11)	(1.74)	(1.77)
Nutrition	17.27	0.31	7.99	0.53	-1.05	-0.26
		(2.94)		(2.96)	(2.67)	(2.43)
Electricity	99.23	0.75	11.28	0.36	-5.43	-2.52
		(0.52)		(0.57)	(10.95)	(5.38)
Sanitation	89.78	-1.85	10.73	-3.61	-62.26***	-32.74***
		(3.84)		(2.97)	(5.52)	(3.50)
Water	30.25	3.06	4.16	2.91	-7.32	-2.20
		(5.86)		(5.48)	(4.94)	(4.78)
Floor	44.95	4.46	6.47	-6.34	1.08	-2.65
		(7.68)		(4.07)	(6.00)	(4.54)
Cooking fuel	99.51	-0.32	11.17	-0.03	0.97	0.47
		(0.50)		(0.59)	(0.68)	(0.58)
Assets	48.03	0.95	5.9	-7.49	-6.37	-6.97
		(4.28)		(5.61)	(5.22)	(4.97)
TOTAL	100	100	100	100	100	100

Note: The first column is the deprivation at each dimension in the control group. The second column is the per cent difference at baseline between the project and the control group. Columns 3 to 5 are per cent DD effects of the project on each deprivation index. Clustered standard errors in parentheses calculated using 500 bootstrap replications. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1% The last column Figures are percent contributions of each deprivation index to the global MPI.

a time. This is equivalent to setting the weight of the left-out indicator to 0. We then redistributed the weight of the left-out indicator in a 'nested' way, that is we reassigned the weight of the missing indicator to the other indicators within its welfare dimension (for example, after excluding nutrition, the weight for child mortality increases from 1/6 to 1/3, while all other weights remain unchanged). With 10 dimensions, this exercise produced 10 new poverty estimates. We also conducted the same exercise leaving out two indicators at a time using the same nested procedure, but avoiding combinations that would result in a removal of an entire welfare dimension (for example we did not leave out years of schooling and school attendance at the same time). For the same reason, we did not extend this exercise to more than two components, because this would lead to the removal of entire welfare dimensions, which is against the rationale for building a multidimensional index.

The results are shown in the charts of Figure 7. Effect sizes at each run are reported in increasing order. Blue dots are statistically significant results, while red dots are not statistically significant. Given the large number of tests, statistical significance was assessed against critical values adjusted by the False Discovery Rate. All estimations show a positive impact of MVP on multidimensional poverty. Note however the large variety of results, ranging from 15 to 5 percentage points when leaving out one indicator, and ranging from 17 to 3 percentage points when leaving two indicators out. In addition, impact was statistically significant only in 60% of cases when leaving out one indicator and only in 50% of cases when leaving out two indicators. These simulations suggest that both the size and the statistical significance of the estimated coefficients are highly sensitive to the removal of one or two indicators.

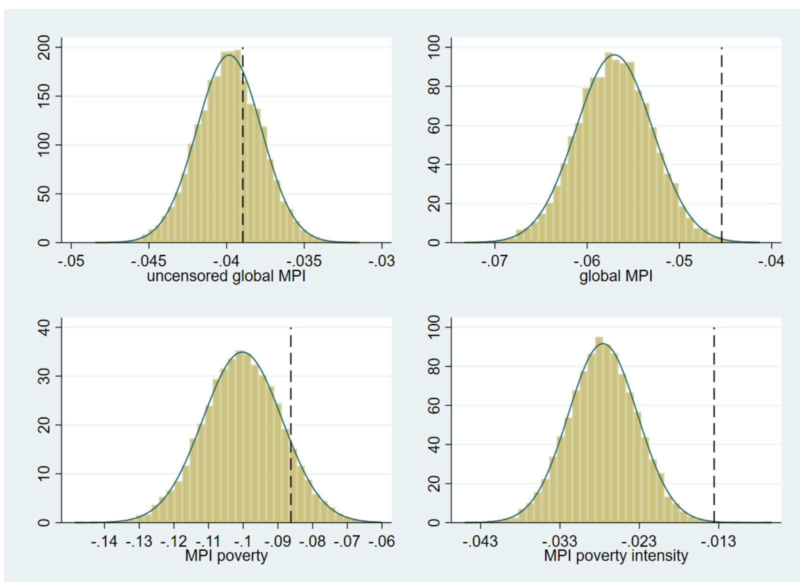
Figure 7: Sensitivity of the global MPI to 1 or 2 indicators



The use of a dual cut-off allows the identification of households suffering joint deprivations. The responsiveness of the index to joint deprivations is one of its most attractive characteristics because it corresponds to the way people normally think about poverty. The global MPI can improve even if the population-level deprivations remain unchanged, for example if the project changes the distribution of deprivations in favour of the most deprived. Could the large impact of MVP on the global MPI be explained by a change in the distribution of deprivations in favour of the most deprived?

To test this hypothesis we randomly shuffled switches in deprivation at the midterm and at the endline, separately for the project and control observations. For example, if 20 households that were deprived in one indicator at the baseline became not deprived at the midterm, we randomly shuffled these changes at the midterm. Similarly, we shuffled the changes occurring in the other direction (from not deprived to deprived). By randomly reallocating the deprivation switches, we preserved the population effect of the intervention but we broke the correlations between switches. The simulations deliver the impact of the intervention produced by uncorrelated switches to which the actual impact of the intervention can be compared.

Figure 8: Simulation of impact of MVP with uncorrelated changes in deprivation



Note: histograms of 10,000 simulated impacts of MVP after reshuffling changes in deprivations. The dashed vertical lines are impacts estimated with the observed data.

The results are shown in Figure 8. The estimated impact of MVP on the global MPI and other indices are compared to simulated impacts in Table 6. As expected, the actual and simulated impacts on the mean deprivation score are very similar and the actual impact is within the 95% normal distribution of the simulated impact. All actual impacts are lower

Table 6: Simulated impacts of uncorrelated changes in deprivations

	Observed impact	Simulated impact	95% interval
Mean deprivation score	-4.28	-3.98	[-4.39 , -3.58]
Global MPI	-4.56	-5.71	[-6.52 , -4.89]
MPI incidence	-8.90	-10.0	[-12.26 , -7.78]
MPI intensity	-1.50	-2.76	[-3.61 , -1.91]

Note: The first column shows the estimated coefficients of the impact of MVP as reported in table 4. The second column reports the average estimated impacts after 10,000 reshuffling of changes in deprivation. The last column is the 95% interval of the simulated normal distribution.

than the simulated ones and outside the 95% normal interval, with the exception of the impact on poverty incidence. The differences indicate that the reductions in deprivations produced by MVP did not reduce the correlations in deprivations among the poor. On the contrary, MVP improved relatively more the conditions of those relatively less deprived, a fact that is in agreement with the lack of observed change in poverty intensity. Hence, the large observed impact of MVP on the global MPI was not a result of a relatively larger improvement in the conditions of households suffering multiple deprivations. On the contrary, the project improved relatively more the conditions of households not suffering from multiple deprivations.

6 Conclusions

The use of the full set of MDG indicators and of a multidimensional poverty index in the evaluation of MVP produced very different results. The project had a limited impact on the MDGs indicators but a substantial impact on the global MPI. What explains the difference?

Firstly, although the global MPI was built to be aligned to the MDG indicators, it includes only a subset of them. It could therefore be expected that the behaviour of the global MPI does not track closely the movement of the MDG indicators. However, the overlap between the global MPI and the MDG indicators is substantial, and the global MPI includes 'core'

MDG indicators, such as child mortality, undernutrion, and school attendance. We would therefore not expect a divergence of results such as the one observed in our data on account of the difference in the included indicators alone.

A second explanation is that the global MPI is sensitive to small changes in deprivations. Because of the 'dual cut-off' method, a household can switch from a poor to a non-poor status simply as a result of a change in one or two deprivations. We showed that MVP particularly improved school attendance and access to sanitation, which accounted together for more than 30% of total deprivation at baseline. The change in the global MPI was therefore driven by small changes in two indicators that do not fully reflect changes in overall living standards in the population.

We conclude with some reflections on the effectiveness of MVP and on the reliability of the global MPI in the evaluation of welfare policies. Was the MVP successful? and should we use the global MPI in the evaluation of development programmes?

The MVP produced modest results, improving just a quarter of the MDG indicators and failing to improve key welfare outcomes.The analysis of the impacts on the MDG indicators shows that the project was able to affect outputs, such as ante-natal care, access to sanitation, and the use of bed nets, but did not affect final outcomes, such as poverty, mortality, and undernutrition. This suggests that important constraints to the realisation of outcomes were not addressed. For example, an increase in the access to sanitation facilities does not imply that the same facilities are used, or a reduction in morbidity and in undernutrition. We tentatively attribute the lack of impact of the intervention to the same factors that led to the failure of similar integrated development programmes in the past (Masset, 2018), namely the selection of project areas with limited growth potential, the use of a top-down approach to project design that ignored relevant contextual enabler and derailers, and the complexity of managing and implementing a large number of disparate and interrelated activities.

Indices have some advantages over multiple indicators in the evaluation of programmes with multiple outcomes. First, they summarise heterogeneous information in a single metric and allow straight comparisons between groups. Second, by summarising impacts on multiple outcomes, they remove the statistical problem of testing multiple hypotheses. Finally, they prevent the selective reporting of results.

Despite these advantages, our conclusion is that the global MPI should be employed with caution in the evaluation of development policies. First, some of these advantages are only apparent. The problem of testing multiple hypotheses can be addressed by using appropriate statistical methods such as, for example, the False Discovery Rate correction. As for the risk of selective reporting, this can be prevented through the use of pre-analysis plans and higher transparency in conducting research.

More importantly, the global MPI is sensitive to changes in few deprivations. It is true that the global MPI can be disaggregated in its components and that the impacts on specific deprivations can be analysed in a transparent way. However, if the global MPI has to be decomposed into its components to be fully understood, then we would prefer to consider a wider set of indicators that are currently not included in the global MPI such as, for example, expenditure poverty, employment, and gender equality.

We are not suggesting however that the global MPI should incorporate all MDG indicators, nor that impact evaluations should build ad-hoc indices based on the expected outcomes of specific programmes. In fact, our use of the global MPI was driven by a desire to prevent the proliferation of indices that are not comparable to each other and whose observed changes have limited practical meaning. The global MPI has the advantage of being simple, transparent and of capturing fundamental dimensions of well-being. The data needed for its construction are minimal and easy to collect. Our analysis found that the global MPI is very sensitive to changes in some deprivations and that it does not include important welfare dimensions affected by development programmes. However, we do not suggest that the global MPI should not be used in evaluation or policy analysis, rather that its use should be further tested and that this might help its refinement.

References

Aaberge, R. and Brandolini, A. (2015). Multidimensional poverty and inequality. In Atkinsons, A. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2, pages 141–216, Amsterdam. Elsevier.

Alkire, S. and Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7-8):476–487.

Alkire, S. and Jahan, S. (2018). The new global MPI 2018: Aligning with the Sustainable Development Goals. *OPHI Working Paper 121, University of Oxford*.

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*, 59:251–274.

Anderson, M. (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*, 103:1481–1495.

Bump, J. B., Clemens, M., Demombynes, G., and Haddad, L. (2012). Concerns about the Millennium Villages Project report. *The Lancet*, 379(May 26):1945.

Clemens, M. and Demombynes, G. (2011). When does rigorous impact evaluation make a difference? The case of the Millennium Villages. *Journal of Development Effectiveness*, 3(3):305–339.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press, New York.

Inbens, G. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences: an Introduction*. Cambridge University Press, New York.

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

Masset, E. (2014). Northern Ghana Millennium Village Impact Evaluation: Analysis Plan. Technical report, ITAD. Available at https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/5903.

Masset, E. (2015). Impact Evaluation of the Sada Northern Ghana Millennium Village Project. Technical report, RIDIE. Available at https://ridie.3ieimpact.org/index.php?r=search/detailView&id=255.

Masset, E. (2018). Integrated development, past and present. *IDS Bulletin*, 49(4):17–32.

Masset, E., Acharya, A., Barnett, C., and Dogbe, T. (2013). An Impact Evaluation Design for the Millennium Villages Project in Northern Ghana. *Journal of Development Effectiveness*, 5(2).

Masset, E., Garcia-Hombrados, J., and Acharya, A. (2020). Aiming High and Falling Low: The SADA-Northern Ghana Millennium Village Project. *Journal of Development Economics*.

Michelson, H. and Tully, k. (2018). The Millennium Villages Project and local land values: Using hedonic pricing methods to evaluate development projects. *World Development*, 101:377–387.

Mitchell, S., Gelman, A., Ross, R., Chen, J., Bari, S., Huynh, U. K., Harris, M. W., Sachs, S. E., Stuart, E. A., Feller, A., Makela, S., Zaslavsky, A. M., McClellan, L., Ohemeng-Dapaah, A., Namakula, P., Palm, C. A., and Sachs, J. D. (2018). The Millennium Villages Project: a retrospective, observational, endline evaluation. *The Lancet*, 6:e500–e513.

Munk, N. (2013). *The Idealist: Jeffrey Sachs and the Quest to End Poverty*. Doubleday, New York.

OPHDI (2018). *Global Multidimensional Poverty Index 2018: The Most Detailed Picture to Date of the World's Poorest People*. University of Oxford, UK.

Pronyk, P. (2012). Errors in a paper on the Millennium Villages Project. *The Lancet*, 379(May 26):1946.

Pronyk, P. M., Muniz, M., Nemser, B., Somers, M. A., McClellan, L., Palm, C. A., Huynh, U. K., Ben Amor, Y., Begashaw, B., McArthur, J. W., Niang, A., Sachs, S. E., Singh, P., Teklehaimanot, A., and Sachs, J. D. (2012). The effect of an integrated multisector model for achieving the Millennium Development Goals and improving child survival in rural sub-Saharan Africa: a non-randomised controlled assessment. *Lancet*, 379(9832):2179–88.

Ravallion, M. (2011). On multidimensional indices of poverty. *Journal of Economic Inequality*, 9:235–248.

Ravallion, M. and Chen, S. H. (2005). Hidden impact? Household saving in response to a poor-area development project. *Journal of Public Economics*, 89(11-12):2183–2204.

Remans, R., Pronyk, P. M., Fanzo, J. C., Chen, J., Palm, C. A., Nemser, B., Muniz, M., Radunsky, A., Abay, A. H., Coulibaly, M., Mensah-Homiah, J., Wagah, M., An, X., Mwaura, C., Quintana, E., Somers, M. A., Sanchez, P. A., Sachs, S. E., McArthur, J. W., and Sachs, J. D. (2011). Multisector intervention to accelerate reductions in child stunting: an observational study from 9 sub-Saharan African countries. *Am J Clin Nutr*, 94(6):1632–42.

Sachs, J., Mcarthur, J., Shmidt-Traub, G., Kruk, M., Bahadur, C., Faye, M., and McCord, G. (2004). Ending Africa's poverty trap. *Brookings Papers on Economic Activity*, 1(2004):117–240.

Sanchez, P., Palm, C., Sachs, J., Denning, G., Flor, R., Harawa, R., Jama, B., Kiflemariam, T., Konecky, B., Kozar, R., Lelerai, E., Malik, A., Modi, V., Mutuo, P., Niang, A., Okoth, H., Place, F., Sachs, S. E., Said, A., Siriri, D., Teklehaimanot, A., Wang, K., Wangila, J., and Zamba, C. (2007). The African Millennium Villages. *Proc Natl Acad Sci U S A*, 104(43):16775–80.

Sen, A. (1992). *Inequality Reexamined*. Harvard University Press, Cambridge.

UNDP (2010). *Human Development Report*. Oxford University Press, New York.

Wanjala, B. and Muradian, R. (2013). Can big push interventions take small-scale farmers out of poverty? Insights from the Sauri Millennium Village in Kenya. *World Development*, 45:147–160.

White, H. (2015). Fieldwork is easy with eyes closed, misunderstanding all you see...and other reflections. *Journal of Development Effectiveness*, 7(4):375–387.

Wilson, J. (2015). The village that turned to gold: a parable of philantrocapitalism. *Development and Change*, 47(1):3–28.

7 Appendix

This appendix includes results further illustrating the estimation of the project effects. Table 7 shows the results of statistical tests of the difference in covariates in the project and the control group before and after applying our matching algorithm. Tables 8, 9 and 10, report the results of impact estimates on the MDG indicators and on the MPI index obtained using a) a simple DiD model without covariates adjustment, and b) a "regression adjustment" model in which the covariates listed in Table 7 and its interaction with the MV indicator were introduced stepwise to the simple DiD model without matching adjustment.

Table 7: Test of covariance balance

Covariate	Unadjusted T-test	Z-value test across strata	F-value test within strata
Household size	-0.83	-0.13	0.71
Age of head	-1.50	0.25	1.42
Education of head	-0.98	-0.24	1.14
Cultivated land	-2.63	-0.31	0.47
Wealth	-1.78	0.01	0.69
Remittances	-4.19	-0.03	2.19
Millet farm	-2.82	-0.10	0.75
Rice farm	3.92	0.09	0.43
Drought shocks	3.02	-0.07	3.75
Flood shock	-3.25	-0.03	1.40
Isolated household	-3.56	-0.03	0.44
Months food insecure	3.62	0.12	0.44
Farmer household	2.52	0.27	1.07
Bank access	-3.13	-0.06	1.02
Metal roof	0.50	-0.09	0.52
Distance to drinking water	-1.49	0.00	0.59
Groundnut farm	-1.95	0.46	1.26

Note: The first column includes t-statistics of tests of the differences in the covariates at baseline. The test in the second column assesses the global balance of each covariate across strata. The test in the third column assesses the balance of each covariate within strata.

Table 8: Impact of MVP on MDG indicators

MDG indicator	Simple Diff-in-diff	Regression adjustment	Sub-classification matching
Proportion of population below $1.25 (PPP) per day	-11.58*	-9.38*	-8.66*
	(2.74)	(2.86)	(2.73)
Proportion of population below the national poverty line	1.88	3.33	1.16
	(3.20)	(3.28)	(3.41)
Poverty gap ratio	0.50	1.42	0.71
	(2.29)	(2.24)	(2.32)
Consumption share of poorest quintile	1.29*	1.24*	0.87
	(0.50)	(0.50)	(0.86)
Employment to population ratio	2.93	2.82	3.06
	(2.48)	(2.40)	(2.41)
Proportion of employed people living below $1 (PPP) per day	-10.42*	-11.06*	-11.04*
	(3.13)	(3.13)	(3.03)
Proportion of own account and contributing family workers in total employment	4.75*	4.70*	4.02*
	(1.68)	(1.68)	(1.93)
Underweight prevalence (children under-5)	-1.04	-0.22	-0.52
	(2.30)	(2.33)	(2.28)
Proportion of population below the food poverty line	1.19	2.62	1.42
	(3.79)	(3.72)	(3.84)
Net attendance ratio in primary education	7.64*	6.54*	7.69*
	(2.89)	(2.79)	(3.16)
Completion rate in primary education	-1.34	-1.33	-1.62
	(2.88)	(2.98)	(3.79)
Young adults (15-24) literacy rate	-4.94	-4.29	-3.36
	(2.88)	(2.84)	(3.33)
Ratio of girls to boys in primary education	-15.96	-17.95*	-18.70
	(8.45)	(8.42)	(9.68)
Share of women employed in the non-agricultural sector	-16.08	-17.08	-8.06
	(10.09)	(8.97)	(13.40)
Under-5 mortality rate	0.58	0.55	0.41
	(2.03)	(2.05)	(2.07)
Infant mortality rate	2.18	2.21	2.02
	(1.85)	(1.85)	(1.89)
Measles immunisation rate (children under-2)	-1.97	-3.10	-4.95
	(4.44)	(3.69)	(3.70)
Proportion of births attended by skilled health personnel	28.82*	25.50*	27.00*
	(4.76)	(4.57)	(4.50)
Contraceptive prevalence rate	7.68*	7.74*	8.22*
	(1.68)	(1.60)	(2.13)
Adolescent birth rate	-4.79	-3.30	-8.67
	(4.75)	(3.86)	(8.31)
Ante-natal care coverage	-3.68	-4.03	-2.94
	(3.53)	(3.48)	(4.06)
Proportion of young adults (15-24) with correct HIV knowledge	0.49	0.15	1.47
	(1.98)	(1.99)	(2.05)
Malaria prevalence (children under-5)	-0.15	-1.20	-5.53
	(4.24)	(4.16)	(4.28)
Proportion of children under-5 sleeping under insecticide treated bed-nets	38.96*	39.25*	39.24*
	(5.38)	(5.37)	(5.85)
Proportion of children under-5 with fever treated with anti-malarial	23.34*	22.22*	16.65
	(7.60)	(7.82)	(8.52)
Proportion of the population using an improved drinking water source	-2.42	-4.42	-0.61
	(3.24)	(3.26)	(3.83)
Proportion of the population using an improved sanitation facility	29.92*	28.93*	30.33*
	(2.91)	(2.80)	(3.16)
Fixed telephone subscriptions rate	0.03	0.03	0.04
	(0.12)	(0.11)	(0.12)
Mobile telephone usage rate	-9.05	-9.10	-7.60
	(5.11)	(5.24)	(5.23)

Note: Table reports estimates of project effects using different estimation models: simple difference-in-difference, adjusted difference-in-difference and sub-classification methods. Coefficients and standard erros were multiplied by 100 to improve readibility as percent differences. Standard errors were calculated using 500 bootstrap replications. A star (*) represents a statistically significant coefficient at 10% with respect to a critical value adjusted by FDR.

Table 9: Impact of MVP on multidimensional poverty

Indicator	Simple Diff-in-diff	Regression adjustment	Sub-classification matching
Multidimensional poverty	-4.36**	-4.23**	-4.56**
index	(1.77)	(1.77)	(1.95)
Multidimensional poverty	-8.35***	-8.21**	-8.90**
incidence	(3.15)	(3.15)	(3.67)
Multidimensional poverty	-1.43	-1.49	-1.50
intensity	(0.99)	(1.01)	(0.95)

Note: Table reports per cent DD effects of the project on each deprivation index using different estimation methods: simple difference-in-difference, adjusted difference-in-difference and sub-classification matching. Standard errors in parentheses calculated using 500 bootstrap replications. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%.

Table 10: Impact of MVP on deprivation indices

Indicator	Simple Diff-in-diff	Regression adjustment	Sub-classification matching
Years of schooling	-4.10	-3.31	-2.30
	(2.78)	(2.85)	(3.01)
School attendance	-6.85**	-6.65**	-7.83**
	(2.74)	(2.70)	(3.40)
Child mortality	0.45	0.31	0.15
	(1.73)	(1.72)	(1.77)
Nutrition	1.79	2.13	-0.26
	(2.72)	(2.68)	(2.43)
Electricity	-2.07	-2.35	-2.52
	(5.51)	(5.47)	(5.38)
Sanitation	-30.95***	-31.23***	-32.74***
	(3.01)	(3.04)	(3.50)
Water	-0.53	-0.62	-2.20
	(3.86)	(3.84)	(4.78)
Floor	-3.56	-4.87	-2.65
	(4.88)	(4.85)	(4.54)
Cooking fuel	0.52	0.52	0.47
	(0.51)	(0.51)	(0.58)
Assets	-7.57*	-7.83*	-6.97
	(4.55)	(4.60)	(4.97)

Note: Table reports per cent DD effects of the project on each deprivation index using different estimation methods: simple difference-in-difference, adjusted difference-in-difference and sub-classification matching. Standard errors in parentheses calculated using 500 bootstrap replications. Stars represent statistical significance levels, whereby * is 10%, ** is 5% and *** is 1%.