The Millennium Villages Project: a retrospective observational end-line evaluation
Supplementary appendix

Shira Mitchell, Andrew Gelman, Rebecca Ross, Joyce Chen, Sehrish Bari, Uyen Kim Huynh, Matthew Harris, Sonia Ehrlich Sachs, Elizabeth A. Stuart, Avi Feller, Susanna Makela, Alan M. Zaslavsky, Lucy McClellan, Seth Ohemeng-Dapaah, Patricia Namakula, Cheryl Palm, Jeffrey D. Sachs

Contents

1 Additional Project Background

The Millennium Villages Project (MVP) was a ten-year, multi-sector, rural development project characterized by:

- implementation of evidence-based, multi-sectoral, and integrated interventions;

- participation of national, regional, and local government and communities in the planning, execution, and/or monitoring of these interventions;
- cost-sharing among the project, government, donors, and local communities; and
- adaptive implementation in response to local conditions.

The MVP simultaneously implemented interventions in agriculture and business, education, health, and infrastructure to address multiple objectives and to enable possible synergistic gains.[1, 2, 3] The ten-year project included two five-year phases. The first phase concentrated on "quick-win" interventions such as:

- free mass distribution of long-lasting insecticide-treated bednets and antimalarial medications;
- elimination of user fees for primary schools and for maternal and child health services;
- expansion of school meals programs;
- construction of roads and other infrastructure; and
- subsidization and provision of improved fertilizers and seeds.

The second phase emphasized increasing government ownership, capacity building, and strengthening of agribusiness, education, and health systems.

See Table 1 for information about each Millennium Village (MV), including its agroecological zone, start date, and end-line data collection period.

Millennium Villages	**Agroecological zone**	**Start date**	**End-line data collection period (2015)**
Potou, Senegal	agro-pastoral millet/sorghum	Q1 2006	July - November
Tiby, Mali	agro-pastoral millet/sorghum	Q1 2006	July - December
Bonsaaso, Ghana	tree crop	Q3 2006	April - October
Pampaida, Nigeria	cereal-root crops mixed (Sudan savanna)	Q2 2006	July - October
Koraro, Ethiopia	highland temperate mixed	Q1 2005	July - October
Ruhiira, Uganda	highland perennial	Q1 2006	May - August
Sauri, Kenya	maize mixed (bimodal)	Q1 2005	June - November
Mayange, Rwanda	highland perennial	Q3 2006	July - September
Mbola, Tanzania	maize mixed (unimodal)	Q2 2006	July - October
Mwandama, Malawi	cereal-root crops mixed (southern miombo)	Q3 2006	March - October

Table 1: Description of the ten Millennium Villages (MVs) covered in this evaluation.

2 Outcomes of Interest

Our outcomes consist of three types: Millennium Development Goal (MDG) indicators;[4] MDG proxy outcomes; and additional project-specified outcomes. We list these outcomes below. "**I**" denotes that the outcome was used in the impact evaluation. "**A**" denotes that the outcome has a target used in assessing target attainment. Proxy outcomes are labeled as approximations to official MDG indicators with a "$\approx$" sign with subscripts.

Targets were defined by the UNDP, unless otherwise indicated: $^{(u)}$ indicates a target defined by UNESCO; $^{(w)}$ indicates a target defined by WHO; and $^{(m)}$ denotes a target defined by the MVP for outcomes without internationally-recognized targets.[5, 6, 7, 8] Targets were either absolute or defined relative to 1990 national-rural data. Where 1990 national-rural data were not available, we used data temporally closest to 1990, see Table 2. Reference data were compiled from a variety of sources, including the World Bank, World Health Organization, Demographic Health Surveys (DHS), and United Nations Statistics Division databases. See Table 3 for site-specific targets.

2.1 Millennium Development Goal indicators and proxies

MDG 1: Eradicate extreme poverty and hunger

1.1 **Proportion of population below 1.25 USD (2005 PPP) per day** (**A**, **I**)

Definition: proportion of people who live below 1.25 USD (2005 PPP) per day, measured as consumption (including consumption from own production)

Target: reduce to 50% of the level in 1990

$\approx_a$ 1.1 **Asset index** (**I**)

Definition: an indicator of household wealth that combines both asset ownership and housing characteristics, reduced to one dimension using principal component analysis (PCA). For a formal definition see equation 3 in Filmer and Pritchett.[9, 10]

Target: no target

1.2 **Poverty gap ratio** (**A**, **I**)

Definition: the Foster-Greer-Thorbecke metric $\mathsf{FGT}_1 = \frac{1}{n}\sum_{i=1}^{n_0}\frac{z-y_i}{z}$ summing over all n_0 people below the poverty line, $z = 1.25$ USD (2005 PPP), where y_i is the consumption of person i, and n is the number of people sampled.[11] (Indicator 1.1 is FGT_0)

Target: reduce to 50% of the level in 1990

1.8 **Proportion of children under five who are moderately or severely underweight** (**A**, **I**)

Definition: proportion of children aged 12 to 59 months with weight-for-age z-score of < -2 based on the WHO standard

Target: reduce to 50% of the level in 1990

$\approx_s$ 1.8 **Proportion of children under five who are moderately or severely stunted** (**A**, **I**)

Definition: proportion of children aged 12 to 59 months with length-for-age or height-for-age z-score of < -2 based on the WHO standard

Target: reduce to 50% of the level in 1990

$\approx_w$ 1.8 **Proportion of children under five who are moderately or severely wasted** (**A**, **I**)

Definition: proportion of children aged 12 to 59 months with weight-for-length or weight-for-height z-score of < -2 based on the WHO standard

Target: reduce to 50% of the level in 1990

MDG 2: Achieve universal primary education

$\approx_n$ 2.1 **Adjusted net attendance ratio in primary education** (**A**, **I**)

Definition: proportion of children of official primary school age (country specific) who attend primary or higher education

Target: $\geqslant 90\%^{(m)}$

$\approx_g$ 2.1 **Gross attendance ratio for primary education** (**A**, **I**)

Definition: total attendants in primary school, regardless of age, expressed as a proportion (which can exceed 100%) of the population of official primary school age (country specific)

Target: $\geqslant 90\%^{(m)}$

2.2 **Proportion of pupils starting first grade who reach last grade of primary education** (**A**, **I**)

Definition: estimated probability of a student in first grade advancing to the end of primary school, subject to retention rates in the year of the survey, estimated by the reconstructed cohort method[4]

Target: $\geqslant 90\%^{(m)}$

MDG 3: Promote gender equality and empower women

3.1 **Gender parity in primary education** (**A**, **I**)

Definition: ratio of girl gross attendance ratio to boy gross attendance ratio

Target: $0.97 - 1.03^{(u)}$

MDG 4: Reduce child mortality

4.1 **Under-five mortality rate** (**A**, **I**)

Definition: estimated probability of a child dying before age five (usually reported as deaths per 1000 live births), subject to survival rates in the five years preceding the survey[12]

Target: reduce to 33% of the level in 1990

4.2 **Infant mortality rate** (**A**, **I**)

Definition: estimated probability of a child dying before age one (usually reported as deaths per 1000 live births), subject to survival rates in the one year preceding the survey[12]

Target: reduce to 33% of the level in 1990

4.3 **Proportion of one-year-old children immunized against measles** (**A**, **I**)

Definition: proportion of children aged 12-23 months who received a measles vaccine before their first birthday

Target: $\geqslant 90\%^{(w)}$

MDG 5: Improve maternal health

5.2 **Proportion of births attended by skilled personnel** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years with a live birth in the last two years who were attended by skilled health personnel during their most recent live birth

Target: reduce proportion of unattended births to 25% of the level in 1990

5.3(A) **Contraceptive prevalence rate, any method** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years who are currently married or in a union where she or her partner is using any contraceptive method

Target: 25% nominal increase from level in $1990^{(m)}$

5.3(M) **Contraceptive prevalence rate, modern method** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years who are currently married or in a union where she or her partner is using a modern contraceptive method

Target: 25% nominal increase from level in $1990^{(m)}$

5.5(1) **Antenatal care coverage, at least one visit with a skilled provider** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years with a live birth in the last two years who received antenatal care at least once with a skilled provider (doctors, nurses, or midwives) during their last pregnancy that resulted in a live birth

Target: $\geqslant 80\%^{(m)}$

5.5(4) **Antenatal care coverage, at least four visits with any provider** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years with a live birth in the last two years who received antenatal care with any provider at least four times during their last pregnancy that resulted in a live birth

Target: $\geqslant 80\%^{(m)}$

MDG 6: Combat HIV/AIDS, malaria and other diseases

$\approx_p$ 6.1 **Proportion of pregnant women tested for HIV during their pregnancy** (**A**, **I**)

Definition: proportion of women aged 15 to 49 years with a live birth in the last two years who received a test for HIV/AIDS during their last pregnancy that resulted in a live birth.

Target: $\geqslant 90\%^{(m)}$

6.3 **Proportion of population aged 15 to 49 years with comprehensive correct knowledge of HIV/AIDS** (**A**, **I**)

Definition: proportion of population aged 15 to 49 years who correctly identify two main ways of preventing sexual transmission of HIV, who reject two common local misconceptions about HIV transmission, and who know that a healthy-looking person can transmit HIV.

Target: $\geqslant 90\%^{(m)}$

$\approx_k$ 6.6 **Proportion of children under five who tested positive for malaria** (**I**)

Definition: proportion of children aged 12 to 59 months who tested positive for malaria

Target: no target

$\approx_s$ 6.6 **Proportion of school-aged children who tested positive for malaria** (**I**)

Definition: proportion of school-aged children (five to 14 years old) who tested positive for malaria

Target: no target

$\approx_w$ 6.6 **Proportion of women who tested positive for malaria** (**I**)

Definition: proportion of women aged 15 to 49 years who tested positive for malaria

Target: no target

$\approx_m$ 6.6 **Proportion of men who tested positive for malaria** (**I**)

Definition: proportion of men aged 15 to 49 years who tested positive for malaria

Target: no target

6.7 **Proportion of children under five who slept under a bednet the night before** (**A**, **I**)

Definition: proportion of children aged 12 to 59 months who slept under an insecticide-treated bednet the night before

Target: $\geqslant 80\%^{(w)}$

$\approx_p$ 6.7 **Proportion of pregnant women who slept under a bednet the night before** (**A**, **I**)

Definition: proportion of pregnant women who slept under an insecticide-treated bednet the night before

Target: $\geqslant 80\%^{(w)}$

$\approx_h$ 6.7 **Proportion of households with at least one bednet** (**A**, **I**)

Definition: proportion of households with at least one insecticide-treated bednet

Target: $\geqslant 90\%^{(w)}$

$\approx_n$ 6.7 **Proportion of people who used a bednet correctly the night before** (**A**, **I**)

Definition: proportion of people who slept under an insecticide-treated bednet with at most one other person the night before. NOTE: the numerator and denominator exclude anyone who slept at someone else's home, even if they slept under an insecticide-treated bednet.

Target: $\geqslant 90\%^{(w)}$

MDG 7: Ensure environmental sustainability

7.8 **Proportion of people who use an improved drinking water source** (**A**, **I**)

Definition: proportion of people who use an improved drinking water source (piped, tap, borehole, protected well, protected spring, rainwater)

Target: reduce proportion without access to 50% of the level in 1990

7.9 **Proportion of people who use an improved sanitation facility** (**A**, **I**)

Definition: proportion of people who use an improved sanitation facility (flush, ventilated improved pit latrine, pit latrine with slab, composting toilet)

Target: reduce proportion without access to 50% of the level in 1990

MDG 8: Develop a global partnership for development

$\approx_a$ 8.15 **Proportion of household that own at least one mobile phone** (**A**, **I**)

Definition: proportion of households that own at least one mobile phone (with or without cellular subscription)

Target: $\geqslant 80\%^{(m)}$

2.2 Millennium Villages Project outcomes

a. Agriculture

a.2 **Proportion of farming households that use mineral fertilizer** (**A**, **I**)

Definition: proportion of farming households (all households that use any land for farming or crop production) that reported using any mineral fertilizer on farms over the past one year

Target: $\geqslant 80\%^{(m)}$

a.4 **Proportion of farming households that use improved seeds** (**A**, **I**)

Definition: proportion of farming households that reported using improved seeds on farm fields over the past one year

Target: $\geqslant 80\%^{(m)}$

b. Education

b.1 **Net attendance ratio for preschool** (**A**, **I**)

Definition: proportion of children of official preschool age (country specific) who attend preschool

Target: $\geqslant 90\%^{(m)}$

b.3 **Net intake rate for the first grade of primary school** (**A**, **I**)

Definition: proportion of children of official primary-school-entrance age (country specific) who enter the first grade of primary education

Target: $\geqslant 90\%^{(m)}$

c. Health

c.1 **Proportion of children under six months who are exclusively breastfed** (**A**, **I**)

Definition: proportion of children under six months who have received only breast milk - no water, other liquids, or foods

Target: $\geqslant 50\%^{(m)}$

c.2k **Proportion of children under five who tested positive for anemia** (**I**)

Definition: proportion of children aged 12 to 59 months whose hemoglobin concentration is below the anemia threshold of 11 g/dl

Target: no target

c.2s **Proportion of school-aged children who tested positive for anemia** (**I**)

Definition: proportion of school-aged children (five to 14 years old) whose hemoglobin concentration is below the anemia threshold of 12 g/dl

Target: no target

c.2w **Proportion of women who tested positive for anemia** (**I**)

Definition: proportion of women aged 15 to 49 years whose hemoglobin concentration is below the anemia threshold of 12 g/dl

Target: no target

c.2m **Proportion of men who tested positive for anemia** (**I**)

Definition: proportion of men aged 15 to 49 years whose hemoglobin concentration is below the anemia threshold of 13 g/dl

Target: no target

2.3 Updates to the outcome list from the evaluation protocol

Here we note any updates (beyond typographical errors) from the outcomes list provided in our evaluation protocol.[13] First, this paper only focuses on analysis of survey data, so all outcomes from operational data, which were collected only in the MVs (not in comparison villages), have been removed from our outcomes list.

Calculation of asset index

Our survey tool asked about ownership of 95 items. The standard method of calculation is to include all 95 assets as binary ownership variables.[9, 10] However, this methodology does not take into account inferior versus superior goods. For example, owning a kerosene lamp is treated the same as owning a light bulb.

Thus, in addition to the standard approach, we calculated the asset index using known orderings to create scores for various goods, which were then fed into a principal component analysis (PCA) to obtain the asset index. Both approaches yielded very similar results, with the scoring approach gaining a small amount of precision relative to the standard approach. For this reason, we present only the scoring approach.

Diet diversity score

The outcome we defined (a daily diet diversity score) was not aligned with our survey tool, which asked about frequency of consumption in a one month recall period. Thus, without strong assumptions, the only quantity that could be calculated is a monthly diet diversity score. We leave the calculation of outcomes from this survey tool to future publications, focusing instead on anthropometric and hemoglobin measurements.

Malaria outcomes with small sample sizes

Due to small sample sizes, we removed three malaria outcomes: (1) among children under five with fever, the proportion who were tested for malaria, (2) among children under five with fever tested for malaria, the proportion who received a positive test result for malaria, (3) among children under five with positive result for malaria, the proportion who took any anti-malarial.

Addition of outcomes from blood testing

We added outcomes from blood testing that were not included in the evaluation protocol: prevalence of malaria and anemia by age-sex groups.

Primary school grade-specific outcomes

We calculated an outcome for the adjusted net attendance ratio in primary education and leave grade-specific measures to other analyses.

Agricultural outcomes

In our evaluation protocol, we had proposed to calculate the average amount of nitrogen used by farming households. However, this outcome does not take into account the size of the farming plots, and it averages households who use zero mineral fertilizer together with households who use mineral fertilizer. Instead, we focused on the proportion of farming households who used mineral fertilizer.

Our evaluation protocol also included an agricultural income outcome; however, due to concerns about the quality of these data, we excluded them from our analysis.

2.4 Targets per MV1

Targets are either absolute or defined relative to 1990 national-rural data. For outcomes with relative targets, each country's 1990 (or temporally closest to 1990) data are shown in Table 2, with corresponding site-specific targets in Table 3.

Outcome	Senegal	Mali	Ghana	Nigeria	Ethiopia	Uganda	Kenya	Rwanda	Tanzania	Malawi
1.1 (WB)	71% (1994)	65% (2001)	64% (1992)	36% (1992)	48% (1995)	60% (1992)	48% (1992)	66% (2000)	41% (1992)	47% (1994)
1.2 (WB)	24% (1994)	21% (2000)	19% (1992)	14% (1992)	13% (1995)	21% (1992)	15% (1994)	24% (2005)	12% (1992)	23% (1998)
1.8 (WN)	27% (1993)	34% (2001)	23% (1999)	38% (1990)	44% (2000)	20% (2001)	21% (1993)	25% (1992)	26% (1992)	26% (1992)
$\approx_s$ 1.8 (WN)	40% (1993)	47% (2001)	35% (1999)	53% (1990)	59% (2000)	46% (2001)	42% (1993)	58% (1992)	51% (1992)	57% (1992)
$\approx_w$ 1.8 (WN)	11% (1993)	14% (2001)	11% (1999)	13% (1990)	13% (2000)	5% (2001)	7% (1993)	5% (1992)	8% (1992)	7% (1992)
4.1 (D)	184 (1992)	273 (1995)	149 (1993)	208 (1990)	192 (2000)	159 (1995)	96 (1993)	163 (1992)	152 (1991)	244 (1992)
4.2 (D)	87 (1992)	145 (1995)	82 (1993)	96 (1990)	115 (2000)	88 (1995)	65 (1993)	90 (1992)	97 (1991)	138 (1992)
5.2 (M,D)	29% (1993)	26% (1995)	30% (1993)	23% (1990)	2% (2000)	32% (1995)	40% (1993)	24% (1992)	34% (1992)	51% (1992)
5.3A (D)	3% (1993)	3% (1996)	15% (1993)	4% (1990)	4% (2000)	12% (1995)	31% (1993)	21% (1992)	8% (1992)	12% (1992)
5.3M (D)	1% (1992)	2% (1995)	7% (1993)	2% (1990)	3% (2000)	5% (1995)	25% (1993)	13% (1992)	5% (1992)	6% (1992)
7.8 (U)	43% (1990)	22% (1990)	37% (1990)	30% (1990)	8% (1990)	39% (1990)	32% (1990)	66% (1990)	44% (1990)	33% (1990)
7.9 (U)	22% (1990)	23% (1990)	4% (1990)	36% (1990)	1% (1990)	40% (1990)	27% (1990)	22% (1990)	23% (1990)	41% (1990)

Table 2: National-rural reference data (and corresponding years of data source) used to set 2015 relative targets for a subset of outcomes in Appendices 2.1 and 2.2. All data are expressed in percentages except for outcomes 4.1 and 4.2, which are expressed as deaths per 1000 live births. Data sources include: (WB) World Bank, (WN) WHO NLiS, (D) DHS, (U) UNSTATS, (M) MICS.[14, 15, 16, 4, 17]

Outcome	Senegal	Mali	Ghana	Nigeria	Ethiopia	Uganda	Kenya	Rwanda	Tanzania	Malawi
1.1	36%	32%	32%	18%	24%	30%	24%	33%	20%	23%
1.2	12%	11%	9%	7%	7%	11%	8%	12%	6%	12%
1.8	14%	17%	11%	19%	22%	10%	11%	12%	13%	13%
$\approx_s$ 1.8	20%	24%	18%	27%	29%	24%	21%	29%	25%	29%
$\approx_w$ 1.8	5%	7%	6%	6%	7%	3%	4%	3%	4%	4%
4.1	61	91	50	69	64	53	32	54	51	81
4.2	29	48	27	32	38	29	22	30	32	46
5.2	18%	18%	18%	19%	24%	17%	15%	19%	17%	12%
5.3A	28%	28%	40%	29%	29%	37%	56%	46%	33%	37%
5.3M	26%	27%	32%	27%	28%	30%	50%	38%	30%	31%
7.8	72%	61%	69%	65%	54%	70%	66%	83%	73%	67%
7.9	61%	62%	52%	68%	51%	70%	64%	61%	62%	71%

Table 3: 2015 targets for a subset of outcomes whose relative targets are based on national-rural data before project start dates. All targets are expressed in percentages except for outcomes 4.1 and 4.2, which are expressed as deaths per 1000 live births.

3 Matching

3.1 Matching variables

Geographic data variables

Below we list geographic variables compiled from geographic information system (GIS) databases and used in the matching procedure described in the main paper.

- GADMv2 and GPWv4 - Administrative boundaries (these do not correspond exactly with the census administrative names).[18, 19]
- PCT_MV1, PCT_MV2, PCT_BUFF - Percent of total grid cell area in the MV1, MV2, and ten kilometer buffer zone around the MV
- Agroecological zones.[20]
  - AEZ_1 - Percent of total grid cell area in the irrigated agroecological zone
  - AEZ_2 - Percent of total grid cell area in the tree crop agroecological zone
  - AEZ_5 - Percent of total grid cell area in the highland perennial agroecological zone
  - AEZ_6 - Percent of total grid cell area in the highland temperate mixed agroecological zone
  - AEZ_7 - Percent of total grid cell area in the root crops agroecological zone
  - AEZ_8 - Percent of total grid cell area in the cereal-root crops mixed agroecological zone
  - AEZ_9 - Percent of total grid cell area in the maize mixed agroecological zone
  - AEZ_11 - Percent of total grid cell area in the agro-pastoral millet/sorghum agroecological zone
  - AEZ_12 - Percent of total grid cell area in the pastoral agroecological zone
  - AEZ_13 - Percent of total grid cell area in the sparse (arid) agroecological zone
  - AEZ_14 - Percent of total grid cell area in the coastal artisanal fishing agroecological zone
  - AEZ_COAST - Percent of total grid cell area that is either water, coastal land, or island and is not captured by the agroecological zone data set.
- Access to major cities.[21]
  - ACCESS_M - Mean (across the grid cell area) travel time in minutes to major cities of more than 100,000 population (1987-2004)
- Soil variables.[22]
  - CEC_M - Mean cation exchange capacity in cmol/kg (1950-2005)
  - CLY_M - Mean percent clay composition of the soil (1950-2005)
  - PH_M - Mean soil pH (1950-2005)
- Vegetation index.[23]
  - EVI_1_M - Mean enhanced vegetation index (2000-2005)
- A temperature index created by averaging standardized versions of these four measures, following the same procedure as for the DHS matching indices (see below):[23]
  - LST_D1_M - Mean moderate-resolution imaging spectroradiometer (MODIS) Land Surface Temperature, day (2002-2005)
  - LST_D2_M - Mean MODIS Land Surface Temperature, day (2005-2010)
  - LST_N1_M - Mean MODIS Land Surface Temperature, night (2002-2005)
  - LST_N2_M - Mean MODIS Land Surface Temperature, night (2005-2010)
- Elevation.[24]
  - ELEV_M - Mean elevation in meters (2008)
  - ELEV_STD - Standard deviation of elevation (2008)
- Population density.[18]

– POPD_M - Mean 2005 population density in persons per square kilometer. Grid cells which fall within the GPWv4 water mask were assigned a value of zero. (1990-2000 data, projected to 2005)

Demographic and Health Surveys variables

We used the following DHS datasets to provide matching variables in seven countries: Senegal 2005, Mali 2006, Ghana 2003, Kenya 2003, Uganda 2006, Rwanda 2005, and Malawi 2004.[16] These surveys were conducted before or in the same year the project started in each country, see Table 1.

Below we list the DHS variables that were used to construct three matching indices: asset wealth (constructed by the DHS), education, and health indices. The education and health indices were created by the following procedure: First, we aggregated variables measured within the household to a household-level variable $x_h^{(k)}$ for each household h and variable k. We then standardized each variable by its mean and standard deviation across all households in the country: $\tilde{x}_h^{(k)} = \frac{x_h^{(k)} - E(x_h^{(k)})}{\sqrt{Var(x_h^{(k)})}}$. Finally, we "reoriented" each variable so that larger values indicate higher economic development. For household h, its education (and health) matching index is the mean of all $\tilde{x}_h^{(k)}$ where k is a variable that belongs to the education (and health) matching index. If a variable was missing for a particular household, it contributed zero to the matching index (here we assumed the data are missing completely at random).

- **Asset wealth index**:
  - the first principal component of a list of assets, e.g. source of water; type of sanitation facility; materials used for housing construction; and ownership of TVs, radios, bicycles, land, livestock
- **Education index**:
  - head of household's education level in years
  - children's school attendance status in the previous year
- **Health index**:
  - household-level variables: ownership of a bednet, presence of a hand-washing station, availability of water at the hand-washing station, availability of soap/cleaning agent at the hand-washing station
  - within-household measurements, aggregated to household level: sought treatment for child's last diarrhea episode, vaccinated child for measles, took malaria medicine during all pregnancies in the last five years, child's length-for-age z-score, slept under bednet the night before (child under five), average number of antenatal care visits during all pregnancies in the last five years, sought treatment for child's last fever, ever heard of AIDS (woman), uses any method of contraception (woman)

Small area estimation

We fit small area models for each of the three matching indices, using geographic data to improve our estimates of wealth, education, and health variables.[25, 26, 27, 28, 29] To account for design variables used in the DHS two-stage cluster sampling, our models include levels for clusters and regions within each country.[30] Furthermore, the models include cluster sampling weights, as recommended in the literature.[31, 32, 33, 34, 35]

We fit Fay Herriot models, where the lowest level of the model is approximated using design-based estimates:[36, 37]

$$\begin{aligned} \hat{\bar{x}}_d &\sim N(\bar{X}_d, \nu_d) \text{ for DHS clusters } d \\ \bar{X}_d &\sim N\left(\mathbf{e}_d^T \boldsymbol{\gamma} + \sum_{l=1}^{3} \beta_l I(w_d \geqslant \zeta_l) + \phi_{r[d]}, \sigma^2_{\text{cluster}}\right) \text{ for DHS clusters } d \\ \phi_r &\sim N\left(0, \sigma^2_\phi\right) \text{ for regions (/state/province) } r \text{ within a country} \end{aligned} \tag{1}$$

where $\hat{\bar{x}}_d$ is the standard design-based estimate of the mean matching index in sampled cluster d and ν_d its sampling variance. The $\mathbf{e}_d$ are the first three principal components of the geographic variables, and w_d is the sampling weight for cluster d from the DHS. We used a “degree-0 spline" with knots ζ_l chosen to be the sampling weights’ quartiles.[32] Where not otherwise specified, our priors are weakly informative.

Conversion of geographic data from grid cell level to DHS circle level

In model 1, the geographic variables act as predictors at the DHS cluster level, which can be geographically identified as a DHS circle (due to displacement for anonymity). This required us to convert the geographic data from grid cells to DHS circles.

Let $\text{overlap}_{c,d}$ be the percent of grid cell c overlapped by DHS circle d. Geographic variables are means across each grid cell area except for the administrative units and the elevation standard deviation. For each such variable VAR, and for each DHS circle d, we computed:

$$\text{VAR}_d = \frac{{}_c\, \text{overlap}_{c,d}\text{VAR}_c}{{}_c\, \text{overlap}_{c,d}}.$$

Then, using a similar procedure for the second moment of the elevation variable, we computed the standard deviation of elevation within DHS circle d as follows:

$$\text{ELEV\_STD}_d^2 = \frac{{}_c\, \text{overlap}_{c,d}(\text{ELEV\_STD}_c^2 + \text{ELEV\_M}_c^2)}{{}_c\, \text{overlap}_{c,d}} - \text{ELEV\_M}_d^2,$$

where ELEV_M_c and ELEV_STD_c are the mean and standard deviation of elevation within grid cell c (see Appendix 3.1).

In model 1, we use administrative units to partially pool across regions. To convert from grid cell to DHS circles, we took the mode across grid cells overlapping the DHS circle:

$$\text{ADMIN}_d = \text{mode}_{\text{overlap}_{c,d}>0}\{\text{ADMIN}_c\}.$$

We then took the first three principal components of the geographic variables (now at the DHS circle level) to use as $\mathbf{e}_d$ in model 1.

Conversion of small area estimates from DHS circle level to grid cell level

After fitting model 1, we had samples of $\overline{X}_d$ from the posterior distribution for each DHS cluster d. We converted to samples from the posterior at the grid cell level by computing, for each grid cell that overlaps at least one DHS circle:

$$\overline{X}_c = \frac{{}_d\, \text{overlap}_{c,d}\overline{X}_d}{{}_d\, \text{overlap}_{c,d}}.$$

3.2 Matching procedure

After restricting to neighboring districts and the MV’s agroecological zone (and for seven countries, to grid cells overlapping DHS buffers), our matching algorithm considered each possible set of five comparison grid cells to determine the set that best matched the treatment grid cells, with “best" defined below. Our search space was restricted to sets with at least two of the five matched comparisons lying within the district. For each set of comparison grid cells, we computed the match’s score, a measure of variable imbalance described below.

After exact-matching, let $N_{\text{in-district}}$ be the number of candidate comparison grid cells in the district containing the MV site and let $N_{\text{out-of-district}}$ be the number in districts neighboring (but not containing) the MV. Thus, the number of possible matches was ${}^{5}_{n_{in}=2} \binom{N_{\text{in-district}}}{n_{in}} * \binom{N_{\text{out-of-district}}}{5-n_{in}}$. If this number was greater than the number that could be considered in 48 hours, we first found the best two within-district matches, followed by the best three matches to complement these. This reduced the search space to ${}^{5}_{n_{in}=2} \binom{N_{\text{in-district}}}{n_{in}} + \binom{N_{\text{out-of-district}}}{5-n_{in}}$. If this reduction was still insufficient to reduce the runtime to 48 hours or less, we limited the search space using a variable considered by subject-matter experts to be highly correlated with the potential outcomes (e.g. the asset wealth index). We restricted comparison grid cells to be within an allowable margin of the mean of this particular variable amongst the treatment cells.

As mentioned above, in Nigeria, Ethiopia, and Tanzania, treatment cells did not overlap with DHS buffers. We therefore did not restrict comparison grid cells to overlap DHS buffers for those three countries. In Kenya and Uganda, the treatment cells overlapped DHS buffers, but Kenya only contained one grid cell that overlapped DHS buffers within the district and agroecological zone, and Uganda contained none. Therefore, for Kenya and Uganda, we selected two or three within-district matches using geographic data alone, but restricted the remaining matches to areas with DHS data.

Imbalance measures

When joint distributions of confounding variables are identical between treatment and comparison groups, the simple difference in outcome means is unbiased for the treatment effect. However, with many variables, estimates of the joint density are subject to the curse of dimensionality.[38, 39] We followed the common procedure of working with lower-dimensional summaries, considering one matching variable at a time.[40] For each variable k, let the (sample) means be $\overline{x}_t^{(k)}$, $\overline{x}_{fc}^{(k)}$, and $\overline{x}_{mc}^{(k)}$ for the treatment cells, the full set of candidate comparison cells, and the matched comparison cells, respectively. Let the standard deviations be $s^{(k)}$, $s_t^{(k)}$, $s_{fc}^{(k)}$, and $s_{mc}^{(k)}$ for all grid cells, the treatment cells, the full set of comparison cells, and the matched comparison cells, respectively. The standardized difference in means is widely recommended to check balance: $\frac{\overline{x}_t - \overline{x}_{mc}}{s^{(k)}}$.[39, 41, 38] We also compared the differences in variance using the logarithm of the ratio of standard deviations between treatment and comparison groups, $\ln \frac{s_t^{(k)}}{s_{fc}^{(k)}}$ before matching, and $\ln \frac{s_t^{(k)}}{s_{mc}^{(k)}}$ after matching.[41]

Since we did not anticipate analyzing the MV1 grid cells separately, we did not examine within-pair statistics.[41] We combined the above scores into an overall match score, first by creating a match score for the standardized difference in means:

$$\text{mean\_score} = \bigoplus_{k \in S_{geo}} \frac{|\overline{x}_t^{(k)} - \overline{x}_{mc}^{(k)}|}{s^{(k)}} + w_{DHS} \bigoplus_{k \in S_{DHS}} \frac{|\overline{x}_t^{(k)} - \overline{x}_{mc}^{(k)}|}{s^{(k)}}, \tag{2}$$

where w_{DHS} is a weight used to increase the influence of DHS variables on the choice of matches. We also created a match score for the differences in variance:

$$\text{var\_score} = \bigoplus_{k \in S_{geo}} \left| \ln \frac{s_t^{(k)}}{s_{mc}^{(k)}} \right| + w_{DHS} \bigoplus_{k \in S_{DHS}} \left| \ln \frac{s_t^{(k)}}{s_{mc}^{(k)}} \right|. \tag{3}$$

We combined these two into a total match score as follows:

$$\text{match\_score} = w_{mean} * \text{mean\_score} + \text{var\_score},$$

where w_{mean} is a weight that favors matching closely on means rather than variances.

Because the DHS variables are much closer to our outcomes of interest (they summarize pre-treatment values of the outcome variables), we set $w_{DHS} = 10$. We set $w_{mean} = 2$ to assign more importance to mean matching as opposed to variance matching.

As mentioned above, not all treatment grid cells overlapped with DHS buffers, requiring modification of the above match scores. For Nigeria, Ethiopia, and Tanzania (whose treatment cells did not overlap any DHS buffers) we dropped the terms that measure the imbalance on DHS variables in equations (2) and (3). For Kenya and Uganda, we also considered matched comparison grid cells that did not overlap DHS buffers. In the above match scores, this missingness in DHS data was handled by computing sample means and standard deviations for available grid cells (a valid approach because DHS clusters were randomly sampled).

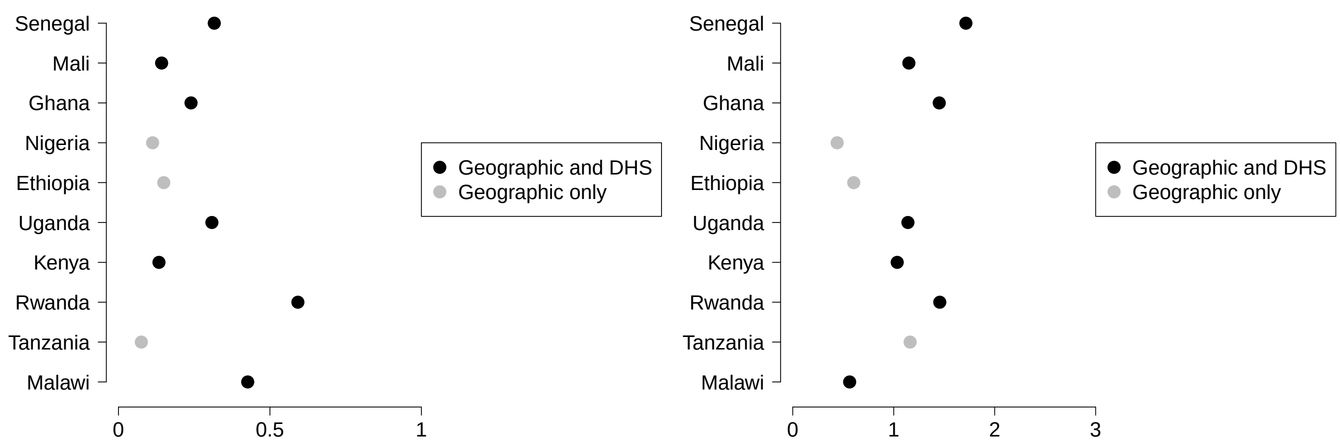
Another complication with the above match scores occurred when either $s_t^{(k)}$ or $s_{mc}^{(k)}$ was zero, making the variance score infinite or undefined. When both $s_t^{(k)} = 0$ and $s_{mc}^{(k)} = 0$, we replaced $\left| \ln \frac{s_t^{(k)}}{s_{mc}^{(k)}} \right|$ with zero, because this represented a good match. When $s_t^{(k)} = 0$ and $s_{mc}^{(k)} \neq 0$, we replaced $\left| \ln \frac{s_t^{(k)}}{s_{mc}^{(k)}} \right|$ with $\left| \ln \frac{\frac{1}{10} s_c^{(k)}}{s_{mc}^{(k)}} \right|$, to penalize large variances in the matched comparison group. When $s_{mc}^{(k)} = 0$ but $s_t^{(k)} \neq 0$, we allowed the match score to be infinite, thereby eliminating these few matches from consideration.

Review by subject-matter experts

The above process included consultation with subject-matter experts to determine whether differences between comparison and treatment were of concern. We presented plots (shown in Figure 2a) to development economists, public health practitioners, geographers, and agricultural scientists. If they voiced concerns about the match on a particular variable, we reran the above algorithm with an adjusted match score that gave more weight to the unbalanced variable.

3.3 Matching results

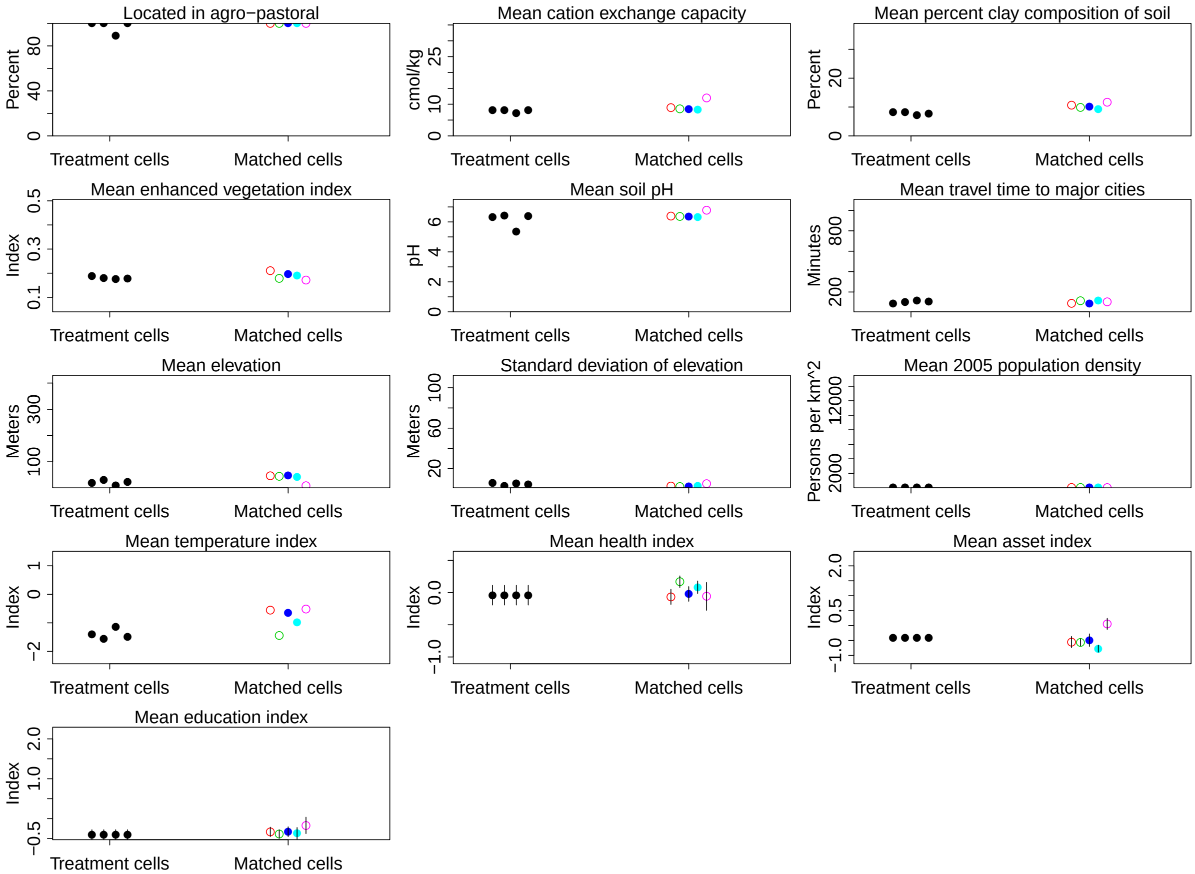
We present the matching results as differences in distributions of each matching variable between treatment and comparison grid cells. We present both scale-free summary measures (see Figure 1) as well as the variables on their original scales (see Figures 2-11). We also present maps of matched grid cells. All matching results are displayed for each country.



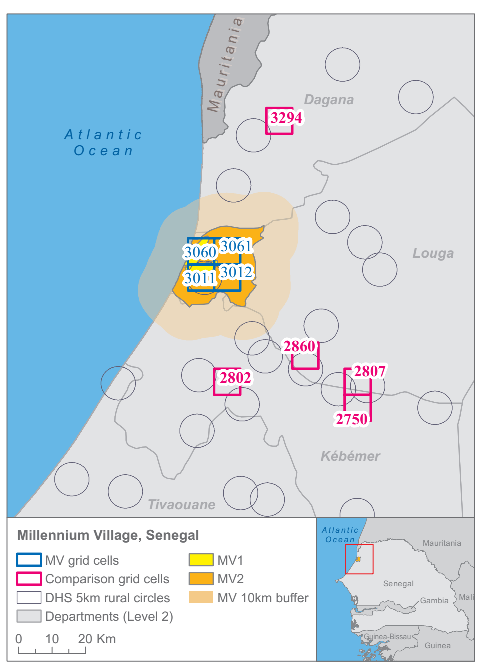
(a) Average standardized difference in means before and after non-exact matching.

(b) Average log ratio of standard deviations before and after non-exact matching.

Figure 1: Summaries of matching results for each country: the average (across matching variables) (a) standardized difference in means, $M_k \frac{|\bar{x}_t^{(k)} - \bar{x}_{mc}^{(k)}|}{s^{(k)}}$*; and (b) log ratio of standard deviations,* $M_k \left| \ln \frac{s_t^{(k)}}{s_{mc}^{(k)}} \right|$*.[39, 41] Matching variables include both geographic variables and the three DHS matching indices (except for Nigeria, Ethiopia, and Tanzania, which only include geographic variables).*

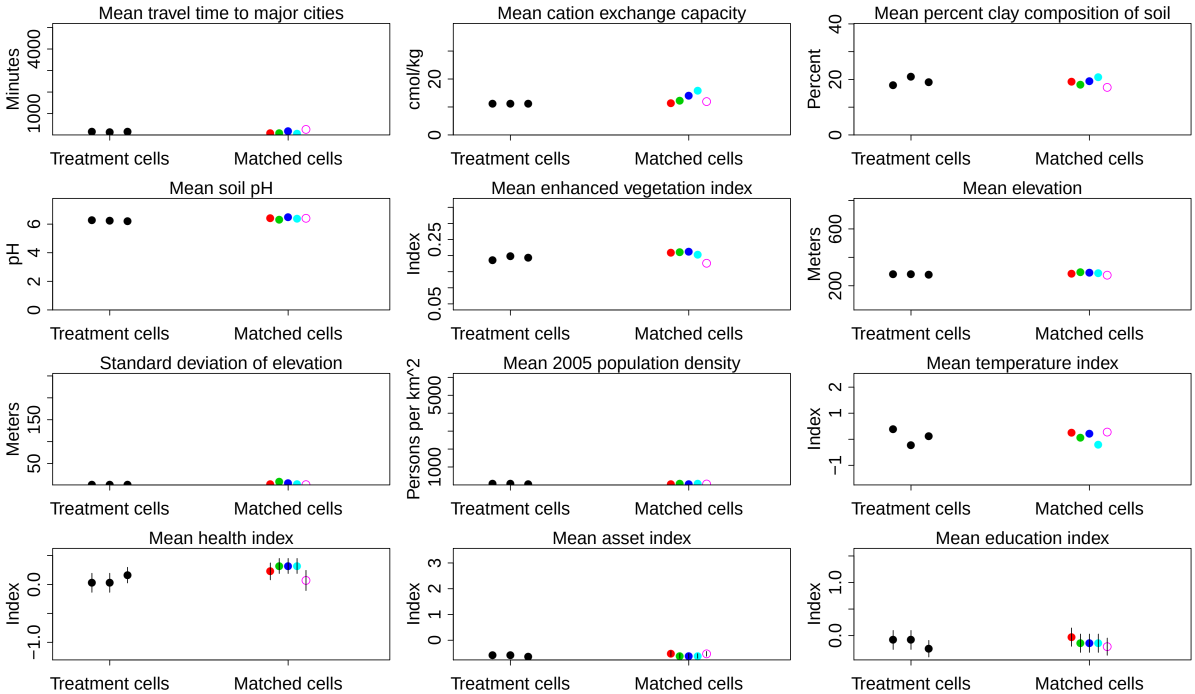


(a) Matching results for Senegal: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 3011, 3012, 3060, 3061 from left to right) and values for matched comparison grid cells are shown in colored circles (2750, 2802, 2807, 2860, 3294). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

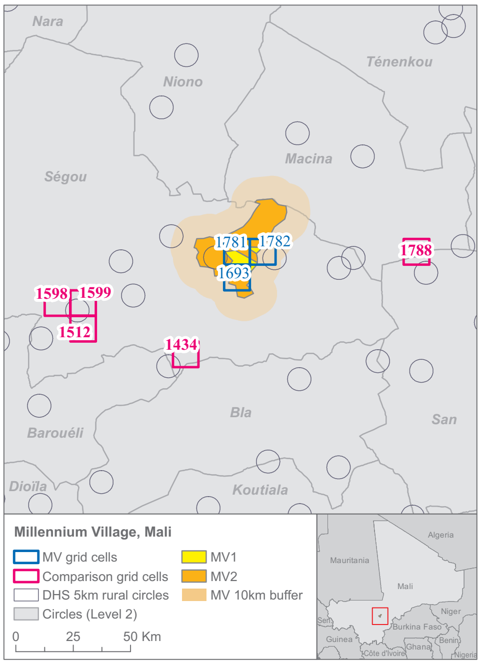


(b) Map of the region surrounding the MV site.

Figure 2: Matching results for Senegal

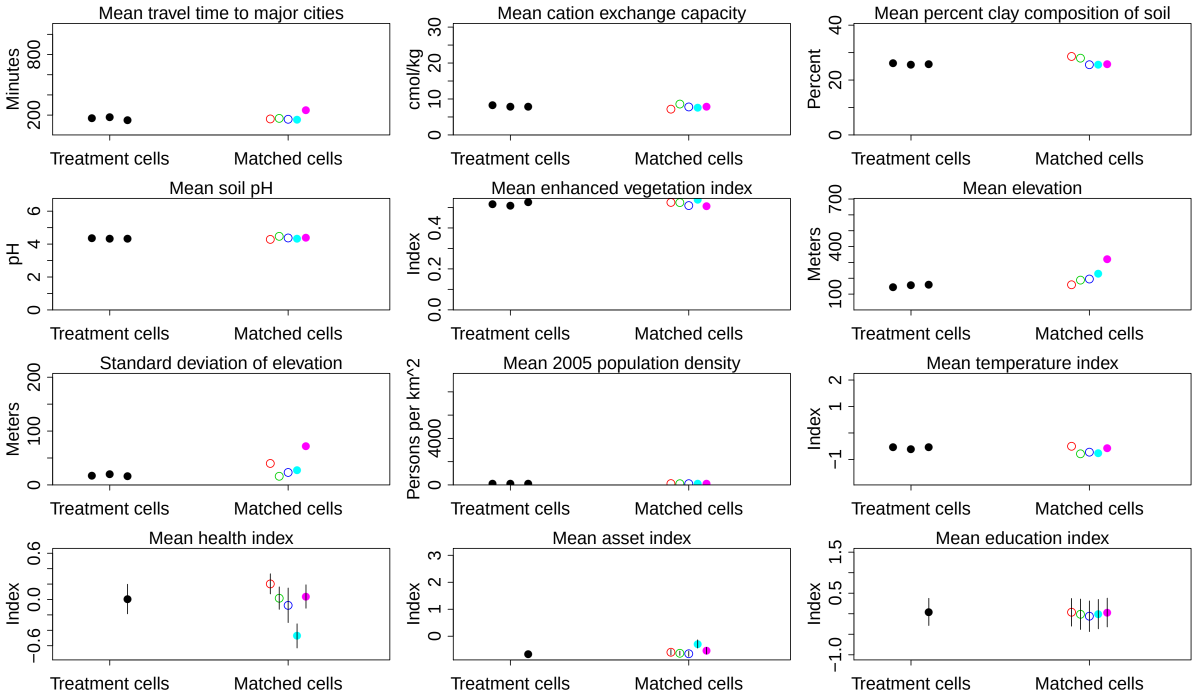


(a) Matching results for Mali: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 1693, 1781, 1782 from left to right) and values for matched comparison grid cells are shown in colored circles (1434, 1512, 1598, 1599, 1788). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

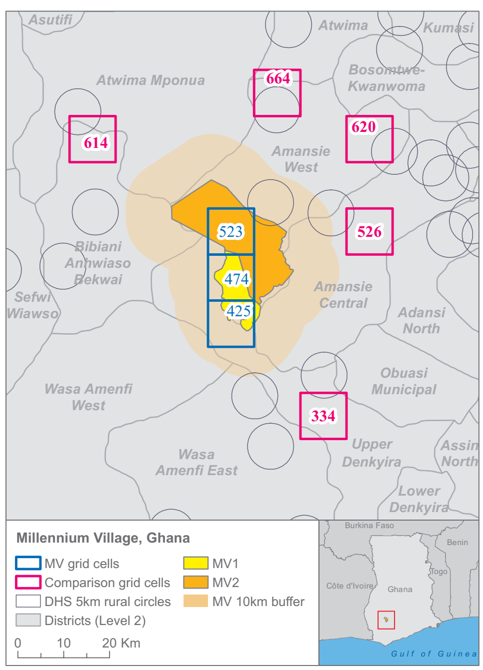


(b) Map of the region surrounding the MV site.

Figure 3: Matching results for Mali

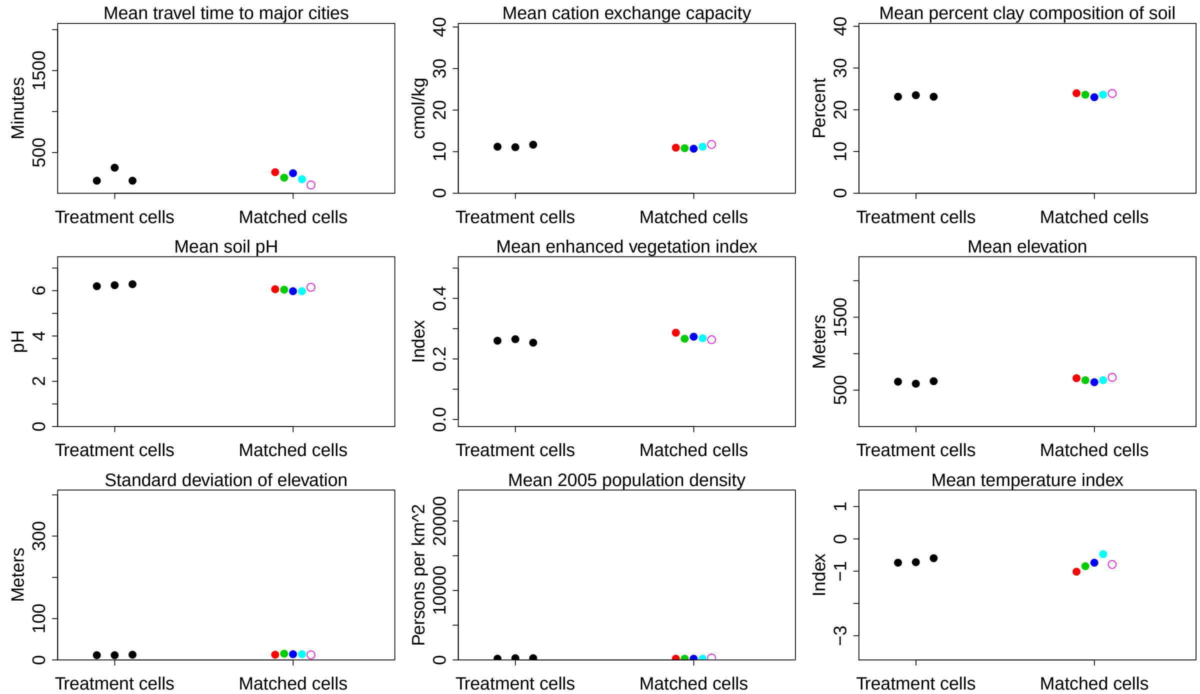


(a) Matching results for Ghana: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 425, 474, 523 from left to right) and values for matched comparison grid cells are shown in colored circles (334, 526, 614, 620, 664). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

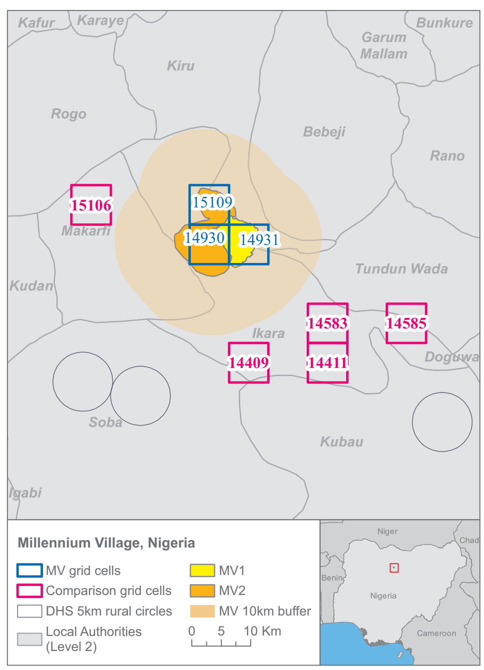


(b) Map of the region surrounding the MV site.

Figure 4: Matching results for Ghana

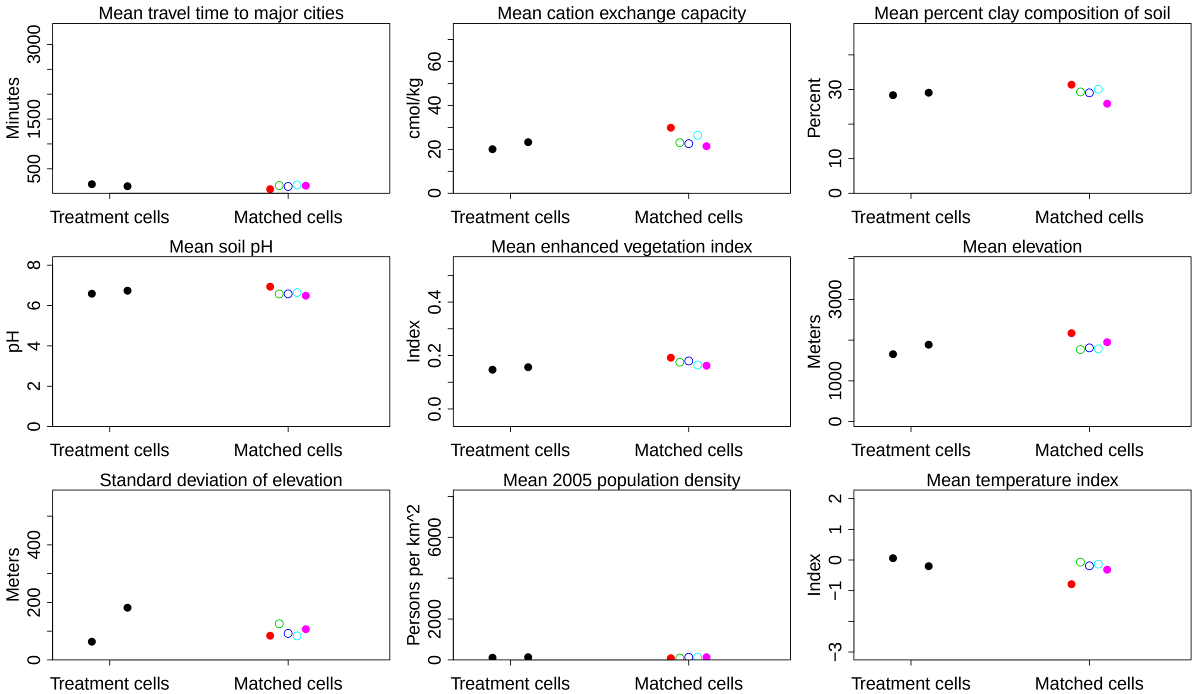


(a) Matching results for Nigeria: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 14930, 14931, 15109 from left to right) and values for matched comparison grid cells are shown in colored circles (14409, 14411, 14583, 14585, 15106). Filled circles represent in-district matches and hollow circles represent out-of-district matches. The axes are scaled from the minimum to maximum value in the country.

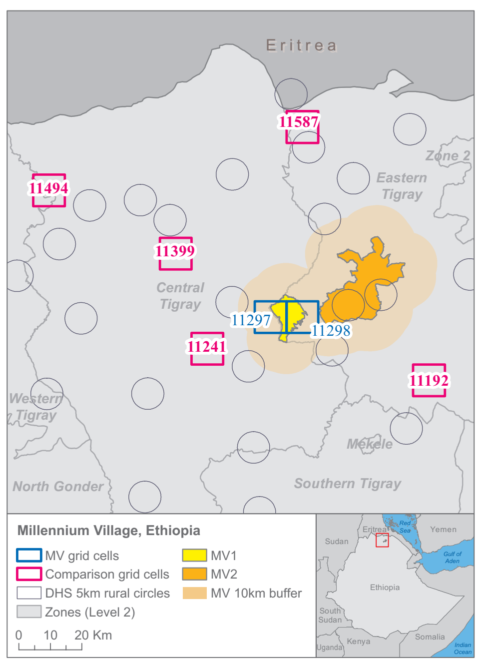


(b) Map of the region surrounding the MV site.

Figure 5: Matching results for Nigeria

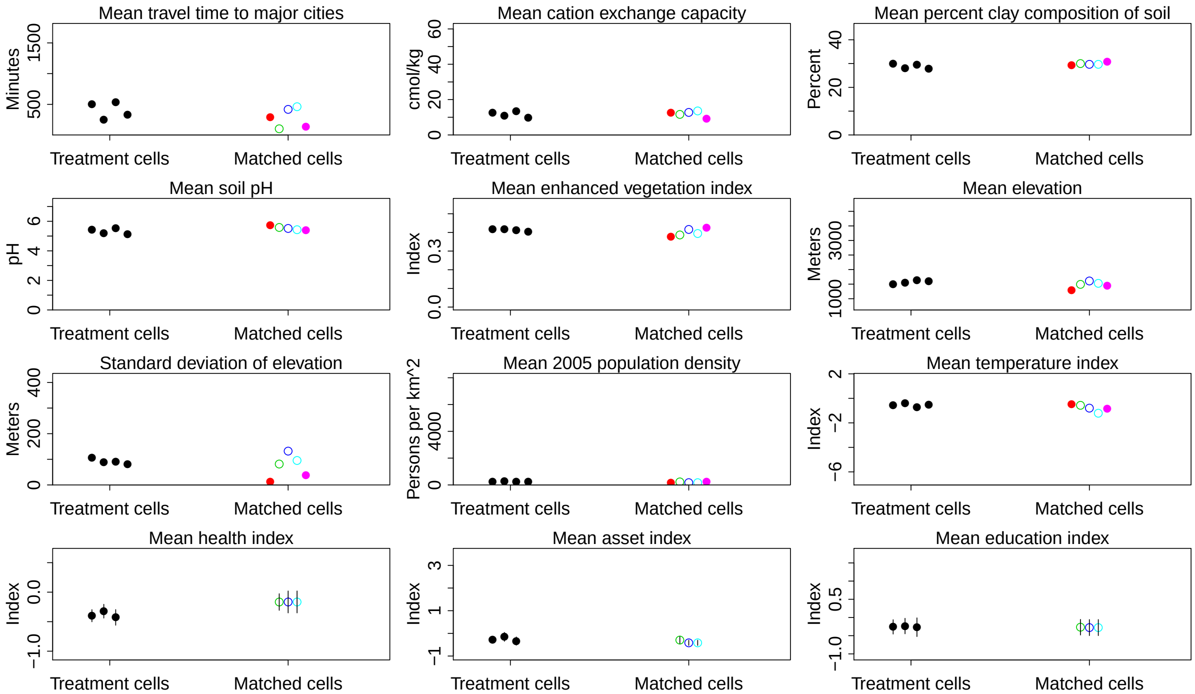


(a) Matching results for Ethiopia: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 11297, 11298 from left to right) and values for matched comparison grid cells are shown in colored circles (11192, 11241, 11399, 11494, 11587). Filled circles represent in-district matches and hollow circles represent out-of-district matches. The axes are scaled from the minimum to maximum value in the country.

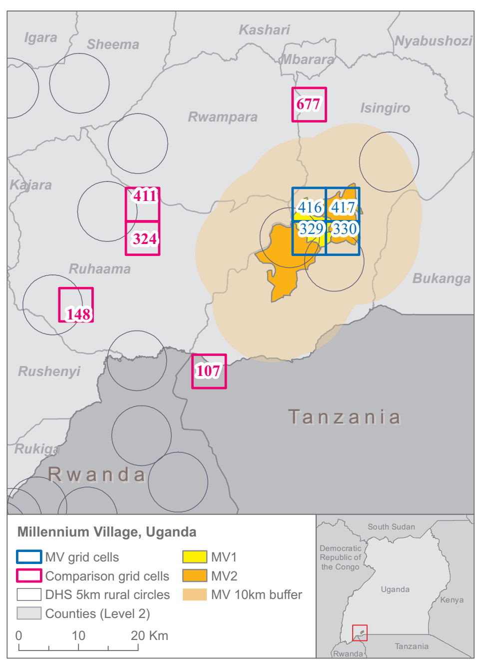


(b) Map of the region surrounding the MV site.

Figure 6: Matching results for Ethiopia

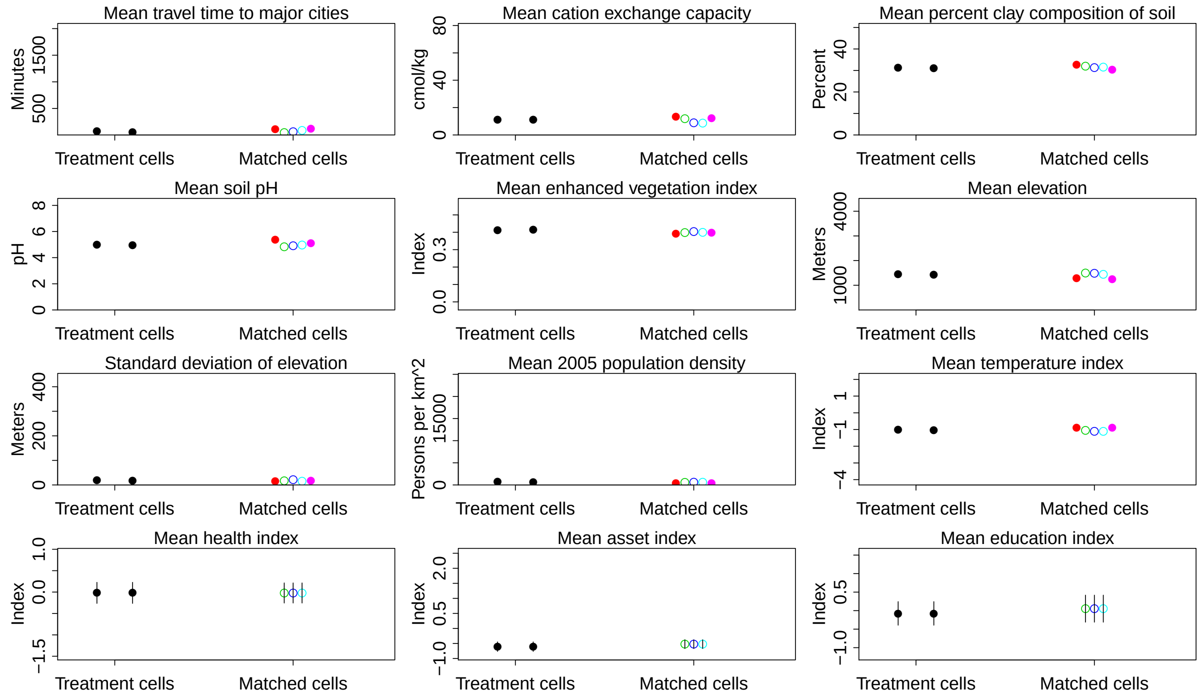


(a) Matching results for Uganda: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 329, 330, 416, 417 from left to right) and values for matched comparison grid cells are shown in colored circles (107, 148, 324, 411, 677). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

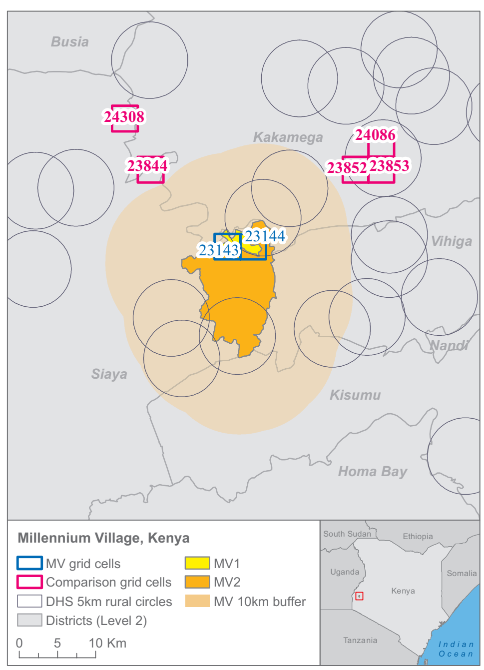


(b) Map of the region surrounding the MV site.

Figure 7: Matching results for Uganda

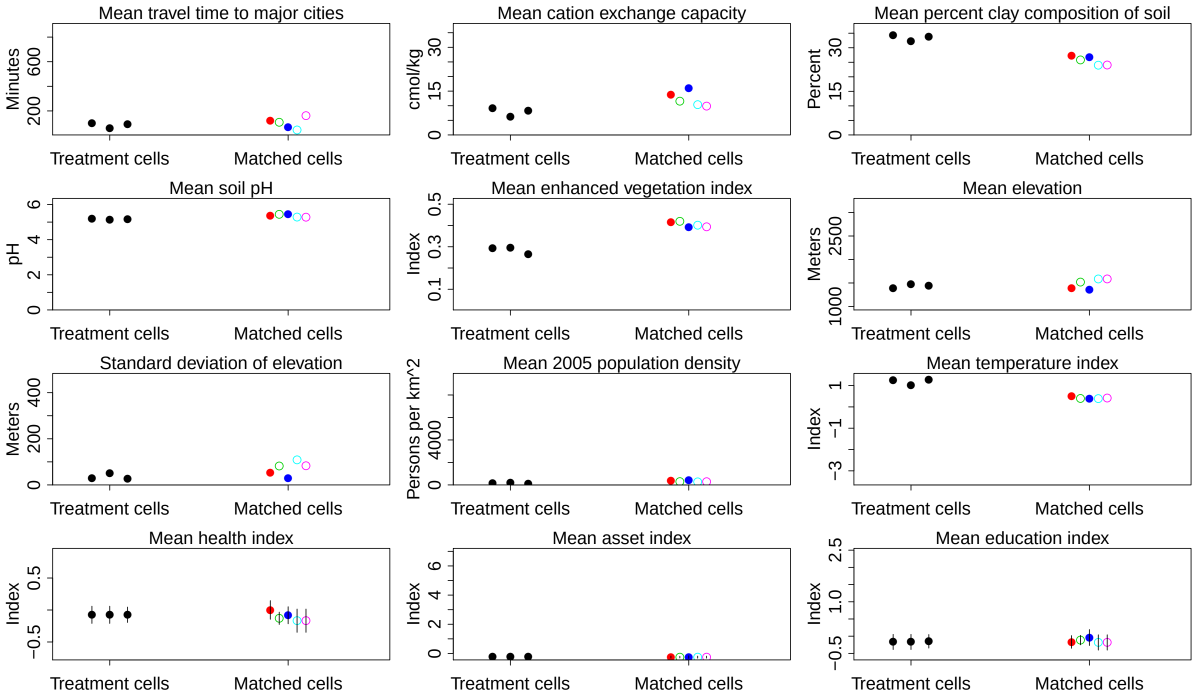


(a) Matching results for Kenya: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 23143, 23144 from left to right) and values for matched comparison grid cells are shown in colored circles (23844, 23852, 23853, 24086, 24308). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

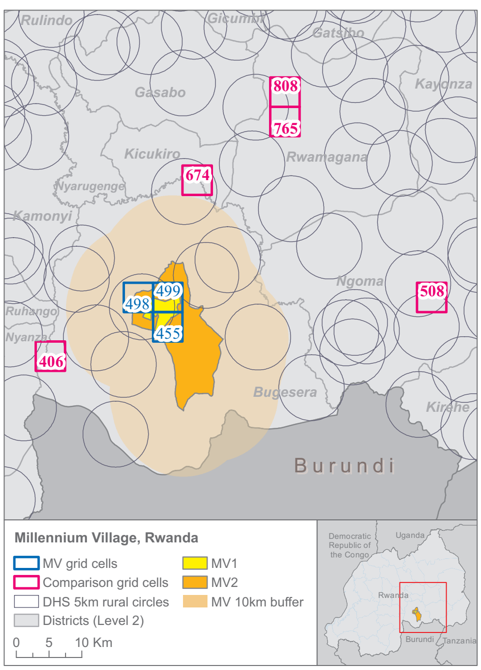


(b) Map of the region surrounding the MV site.

Figure 8: Matching results for Kenya

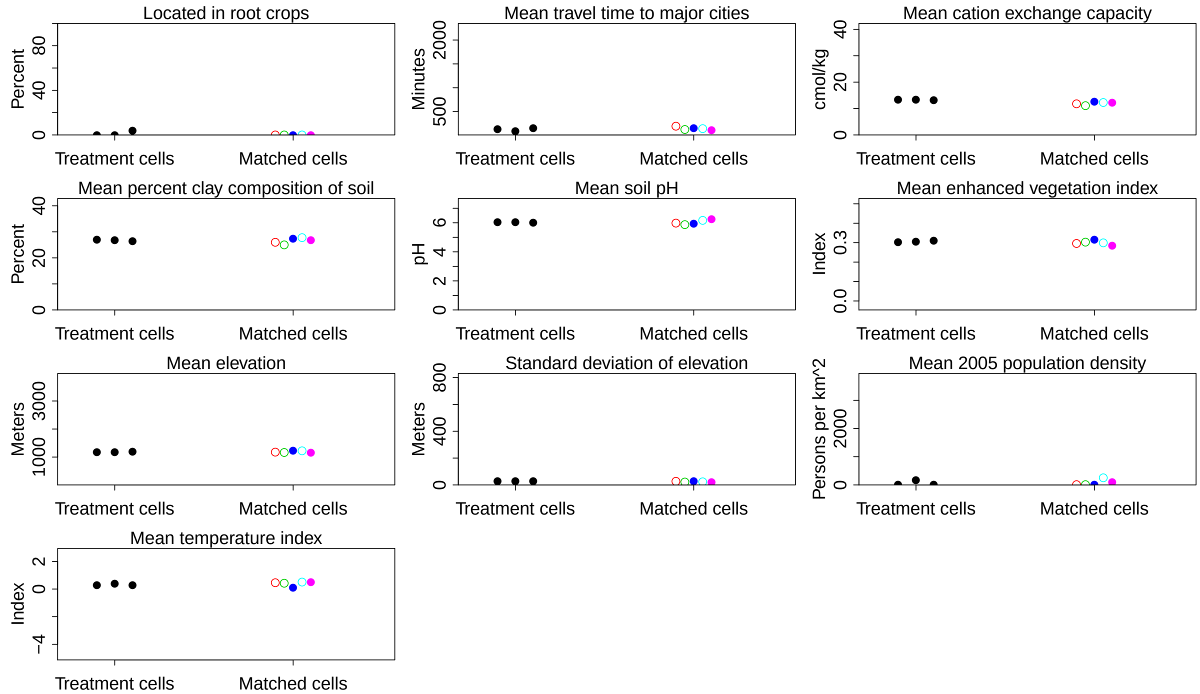


(a) Matching results for Rwanda: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 455, 498, 499 from left to right) and values for matched comparison grid cells are shown in colored circles (406, 508, 674, 765, 808). Filled circles represent in-district matches and hollow circles represent out-of-district matches. Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.

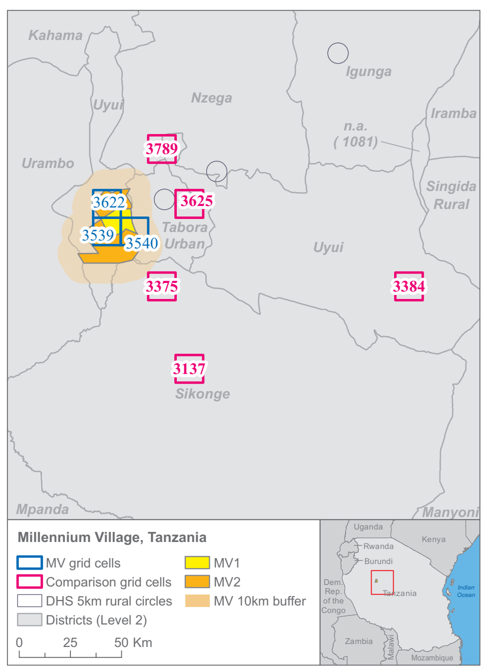


(b) Map of the region surrounding the MV site.

Figure 9: Matching results for Rwanda

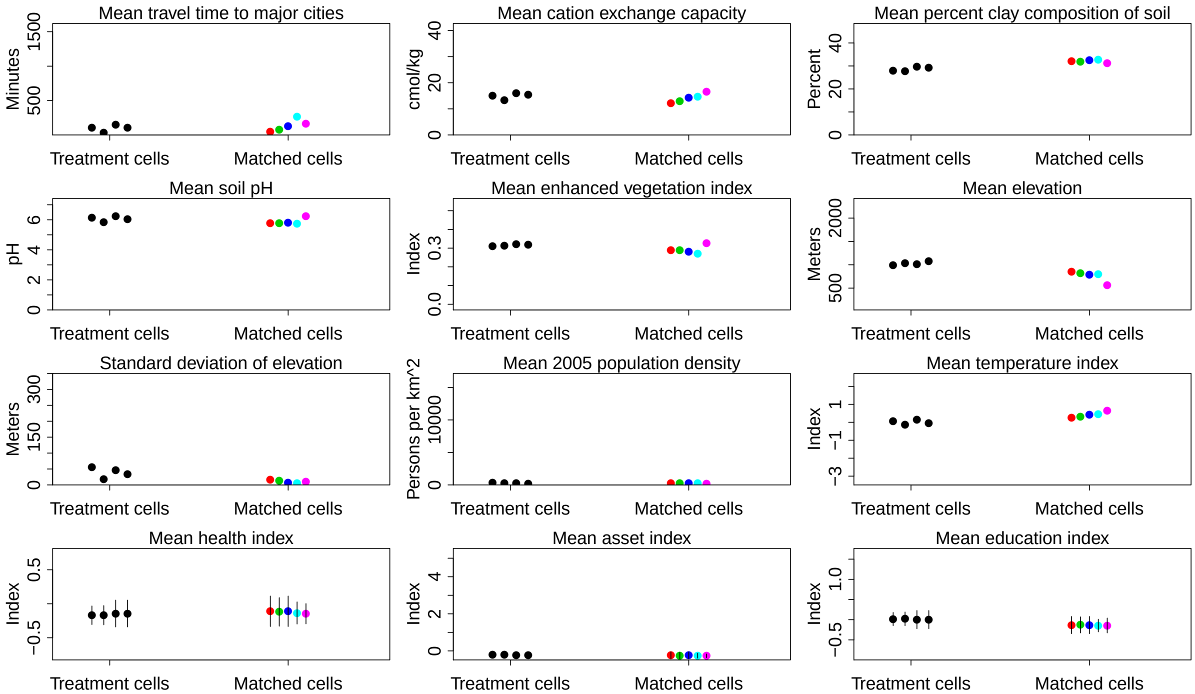


(a) Matching results for Tanzania: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 3539, 3540, 3622 from left to right) and values for matched comparison grid cells are shown in colored circles (3137, 3375, 3384, 3625, 3789). Filled circles represent in-district matches and hollow circles represent out-of-district matches. The axes are scaled from the minimum to maximum value in the country.

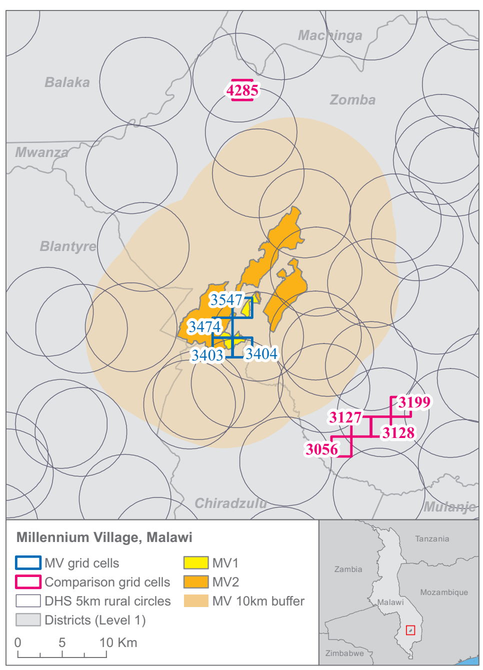


(b) Map of the region surrounding the MV site.

Figure 10: Matching results for Tanzania



(a) Matching results for Malawi: for each matching variable, values for treatment grid cells are shown in black circles (corresponding to grid cells 3403, 3404, 3474, 3547 from left to right) and values for matched comparison grid cells are shown in colored circles (3056, 3127, 3128, 3199, 4285). Where DHS data were available, we present 95% uncertainty intervals for the matching indices from our small area estimation procedure. The axes are scaled from the minimum to maximum value in the country.



(b) Map of the region surrounding the MV site.

Figure 11: Matching results for Malawi

4 Migration

We estimated the extent of migration in and out of the treatment sites, showing that household heads in the MV1s had lived there for almost ten years (on average) at end-line, and household heads in the comparison villages had not lived many years in the MV sites (on average).

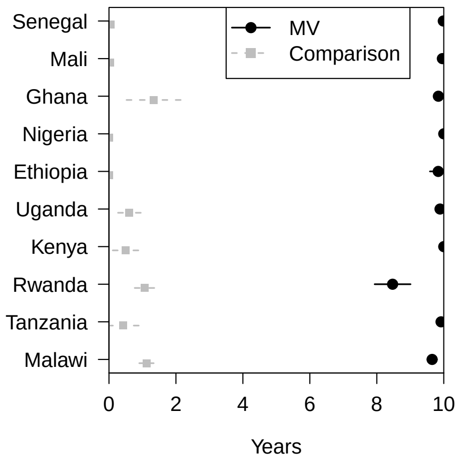


Figure 12: Classical estimates and 95% intervals of uncertainty for the average number of years lived in the MV site since 2005 for each country, based on household survey of migration patterns.

5 Outcome Data Collection

5.1 Data collection procedures

Definition of study population

The study population of interest are households and their members (see definition below) living in the MV1s or comparison villages at the time of the survey. This excludes any households that are abandoned or destroyed and people who died or moved out between the construction of the sampling frame and the survey administration. It also excludes households or members who were reported to be away for the entire duration of the data collection period, without a planned return date.

Definition of household members

We define household members to be those who have lived in the household for at least three of the past 12 months and who 'normally eat from the same pot.' Additionally, the following persons are always considered to be household members: the main provider for the household and infants who are less than three months old.

Household listing and census

As in previous years, prior to survey administration, a census was conducted in each MV1 to record basic demographic data of all household members. To conserve resources, we did not conduct such a census in comparison villages. Rather, prior to sampling households within each comparison village, we created a household list: a list of all non-abandoned households (determined by outside appearance, without consultation of household members) with GPS coordinates identifying the location. From this list, we chose a simple random sample of households. Next, we administered a demographic census in those sampled households, which served as the sampling frame for the within-household sampling.

Data collection and entry

Enumerators were provided with pre-populated survey tools identifying the sampled individuals prior to administration of each survey module. If a sampled individual was not found at home when the enumerator visited, present household members were asked if the sampled person resides in the household and why they were absent. If the person still resided in the household, the enumerators made up to six visits to the household to attempt to reach them. If the person no longer resided in the household, the enumerators were instructed not to follow-up. Generally, they consulted with other household members to determine a time when the person would be available. If the person could not be reached after six attempts, this was recorded as missing data.

Field supervisors randomly conducted spot checks during data collection and all questionnaires were checked after enumeration. Data entry clerks double entered data in CSPro templates with logic checks to minimize errors.[42]

5.2 Survey modules

Sample sizes for each module and age-sex group were determined based on budget, logistics, and relative importance of vulnerable populations.

Household surveys

A survey was administered to household heads (or other knowledgeable members) within the sampled households, capturing data on poverty, agriculture, education, malaria, and water and sanitation. We used the household head's duration of residence in the MV1 as a rough measure of in-migration, but had no feasible way to measure out-migration.

Adult surveys

A sex-specific adult survey was administered to men and women of reproductive age (15 to 49 years) within sampled households. In each MV1, if the total number of men (respectively, women) in sampled households was at most 500, then all were sampled. Otherwise, we took an equal-probability systematic sample of 500 men (women), where households were ordered randomly and people within households were ordered randomly (conceptually similar to stratifying on household).[43] Sampling in each comparison village was the same, with the sample size divided amongst the comparison villages within a country.

The adult male and female surveys included questions on marital status, sexual and reproductive health, and HIV knowledge. The female survey also included questions on birth history and contraceptive use, as well as child health, immunization history, and feeding practices. To get better estimates of child mortality rates, we administered the birth history module to more women than sampled for the adult female survey. We took a simple random sample of extra households to reach 1000 women in each country-treatment group.

Blood and anthropometric data

In each country-treatment group, we sampled four age-sex groups in sampled households for malaria and anemia testing: up to 300 children aged 12 to 59 months, 100 children aged five to 14 years, 100 men aged 15 to 49 years, and 100 women aged 15 to 49 years. In each country-treatment group, we measured weight, height, length, and mid-upper-arm circumference (MUAC) of up to 400 children aged 12 to 59 months. The sampling schemes were as described above for the adult survey.

Anemia was tested using HemoCue HB 301 point-of-care device.[44] Those found to be anemic were referred to the nearest health center. Malaria parasitemia was tested using Rapid Diagnostic Tests (RDTs) from Access Bio Inc.[45] CareStart^{TM} Malaria HRP2/pLDH (Pf/Pv) Combo test G0161 was used in Ethiopia and Rwanda, where *Plasmodium Vivax* is prevalent.[46] In all other sites, CareStart^{TM} Malaria HRP2/pLDH(Pf) test G0181 was used. Both G0161 and G0181 were tested in the WHO-Foundation for Innovative New Diagnostics (FIND) Malaria RDT Evaluation Program and met the WHO recommended selection criteria.[47, 48] Those who tested positive for malaria were treated with Artemisinin-based Combination Therapies.

Weight was measured using Seca 874 weighing scales in all sites except Senegal, where HealthOMeter 498K scales were used. Length and height were measured using portable measuring boards.[49] Recumbent length was measured for children 12 to 23 months of age, and standing height was measured for children 24 to 59 months of age. If bilateral edema was present in a child's feet or if the MUAC measurement for a child (12 months and older) was less than 125 mm, then the child was referred to the nearest health facility for malnutrition assessment and treatment.

Adjusting the sampling frame

To account for the lag between creation of the sampling frame and enumeration, we adjusted the under-five age group to include children aged 12 to 59 months (rather than six to 59 months).

5.3 Missing data

We were unable to collect demographic data from less than 1% of households in the study population, so our analysis excluded these few cases. We imputed the remaining missing data. Before imputation, we combined raw survey data into variables, e.g. computing the weight-for-age z-score from measures of weight and age. In Figure 13 we display the percent nonresponse for variables used to calculate our outcomes. For all but the birth history variables, we imputed missing values using the `mi` package in `R`, which iteratively draws imputed values from the conditional distribution for each variable given the observed and imputed values for other variables.[50, 51] Our procedure took into consideration the survey skip patterns and different data types. Missing birth months were imputed to be July (affecting the cohort definition, not the survival times). The records of less than 2% of children with missing ages of death were dropped.

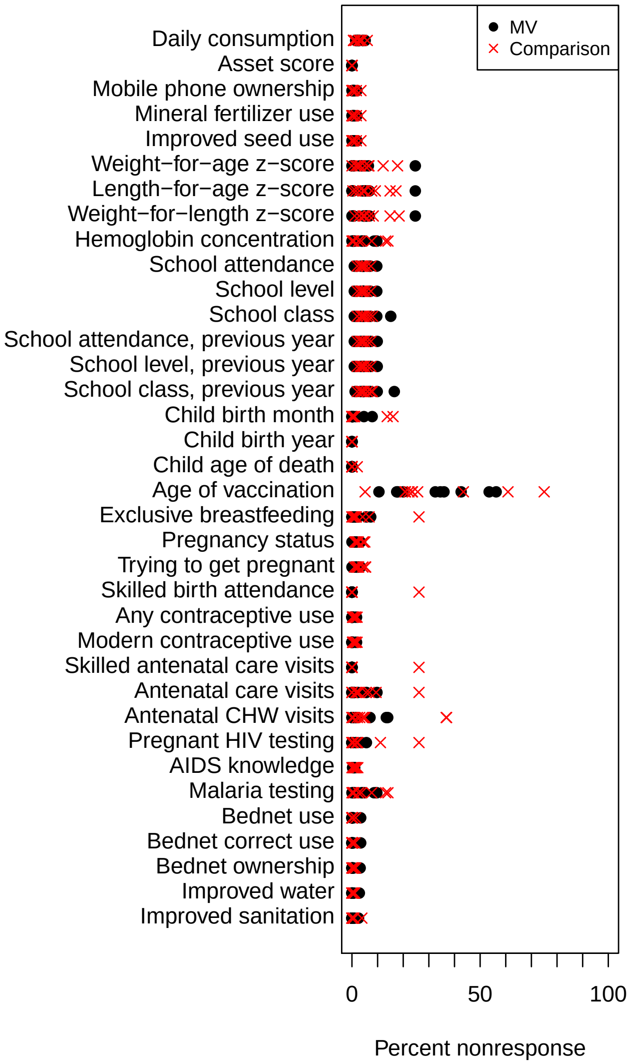


Figure 13: The percent nonresponse in survey variables used to calculate our outcomes of interest, presented for each country-treatment group.

Variable	MV Number sampled	MV Number responded	MV Percent nonresponse	Comparison Number sampled	Comparison Number responded	Comparison Percent nonresponse
Daily consumption	294	285	3%	285	277	3%
Asset score	294	293	0%	285	284	0%
Mobile phone ownership	294	291	1%	285	282	1%
Mineral fertilizer use	294	291	1%	285	282	1%
Improved seed use	294	291	1%	285	282	1%
Weight-for-age z-score	178	170	5%	194	183	6%
Length-for-age z-score	178	169	6%	194	181	7%
Weight-for-length z-score	178	169	6%	194	181	7%
Hemoglobin concentration	411	394	4%	438	419	4%
School attendance	891	854	5%	853	819	4%
School level	891	854	5%	853	819	4%
School class	891	849	6%	853	816	5%
School attendance, previous year	891	854	5%	853	818	4%
School level, previous year	891	854	5%	853	818	4%
School class, previous year	891	848	6%	853	815	5%
Child birth month	1001	987	2%	859	822	3%
Child birth year	1001	1001	0%	859	859	0%
Child age of death	80	80	0%	86	86	0%
Age of vaccination	46	31	32%	51	34	32%
Exclusive breastfeeding	94	92	2%	111	108	4%
Pregnancy status	332	326	2%	341	332	3%
Trying to get pregnant	332	326	2%	341	331	3%
Skilled birth attendance	94	94	0%	111	109	3%
Any contraceptive use	267	266	0%	265	263	1%
Modern contraceptive use	267	266	0%	265	263	1%
Skilled antenatal care visits	94	94	0%	111	109	3%
Antenatal care visits	94	92	2%	111	106	5%
Antenatal CHW visits	94	91	4%	111	102	9%
Pregnant HIV testing	94	93	1%	111	107	4%
AIDS knowledge	629	625	1%	631	625	1%
Malaria testing	411	394	4%	438	419	4%
Bednet use	1608	1603	0%	1544	1541	0%
Bednet correct use	1608	1603	0%	1544	1541	0%
Bednet ownership	294	292	0%	285	283	0%
Improved water	294	292	0%	285	284	0%
Improved sanitation	294	292	0%	285	283	1%

Table 4: Sample sizes, number of responses, and nonresponse percents, by treatment group, averaged across countries.

5.4 Spending data

We documented annual on-site spending by site, stakeholder (project, government, donors, and local community), and sector, within the MV1s as well as the MV2s.[52, 53, 54, 55, 56] We measured spending on activities such as malaria control or irrigation development, not the spending on specific outcomes. No analogous data collection effort was conducted in the comparison villages.

Project spending was reported quarterly via an internal financial tracking system. External stakeholder spending was reported via structured interviews and data collection templates. Community, donor, and government spending data were collected through 2014; internal project spending data were collected through 2015.

6 Statistical Analysis

6.1 Estimation of village-level outcomes

Here we describe the procedure used to obtain estimates and variances for village-level outcomes: $y_{j,t}^{(k)}$ and its estimated variance $V_{j,t}^{(k)}$ for outcome k in village j. The estimation procedure differed by outcome type.

For estimation at the village level, we considered households as the primary sampling units (PSUs). Let s_I be the set of sampled PSUs, $n_I = |s_I|$, and s_h be the sample in PSU h. Given our use of equal-probability sampling and imputation to handle nonresponse, our estimation did not employ survey weights.

Ratio estimation

Most outcomes are the average of some variable y_i (e.g. received HIV test) among individuals or households with $d_i = 1$ (e.g. had a live birth in the last two years). Let $t^{\text{num}} = {}_{i}\, y_i d_i$ and $t^{\text{den}} = {}_{i}\, d_i$ be population totals, with the quantity of interest $R = \frac{t^{\text{num}}}{t^{\text{den}}}$.

Let $y_h = {}_{i \in s_h} y_i d_i$ and $d_h = {}_{i \in s_h} d_i$ be the sample totals in PSU h. We estimated the ratio as $\hat{R} = \frac{{}_{h \in s_I} y_h}{{}_{h \in s_I} d_h}$ and its variance assuming with-replacement sampling of PSUs:[43, 57]

$$\hat{V}(\hat{R}) = \frac{1}{({}_{h \in s_I} d_h)^2} \frac{n_I}{n_I - 1} {}_{h \in s_I} (y_h - \hat{R} d_h)^2,$$

matching formulas used by the R survey package function `svyratio()`.[51, 58]

Taylor linearization

The above ratio variance estimation is a special case of the more general technique of Taylor linearization for variance estimation.[43] We used this more general technique for outcomes 2.2 (proportion of pupils starting first grade who reach last grade of primary education) and 3.1 (gender parity in primary education), which cannot be expressed as ratios of two population totals.

Gender parity in primary education

Define $y_{1i} = 1$ if child i attends primary school and is female (0 otherwise), $y_{2i} = 1$ if child i is of primary school age and is female, $y_{3i} = 1$ if child i attends primary school and is male, $y_{4i} = 1$ if child i is of primary school age and is male. Define totals as $t_g = {}_{i}\, y_{gi}$ for $g = 1, ..., 4$ and our quantity of interest as $\frac{t_1 t_4}{t_2 t_3}$.

Proportion of pupils starting first grade who reach last grade of primary education

Let G = number of grades of primary education. We defined a survival time T = the highest grade that was started, where if $T > 1$, we required that grades one through $T - 1$ were completed in order in $T - 1$ years (i.e. no leaving, repeating, or skipping grades). Thus, the quantity of interest could be expressed as:

$$P(T \geqslant G | T \geqslant 1) = P(T \geqslant 2 | T \geqslant 1) ... P(T \geqslant G | T \geqslant G - 1).$$

Given the data (survey data in year t asking about the previous year $t-1$), we estimated the above probabilities, for example:

$$P(T \geqslant 6 | T \geqslant 5) \approx \frac{\text{number of children in grade 6 in year } t \text{ and in grade 5 in year } t-1}{\text{number of children in grade 5 in year } t-1}.$$

The denominator does not exactly represent those with survival times $T \geqslant 5$, since we did not ask about their completion of grades one through four. However, conditional on $T \geqslant 5$, we know whether $T \geqslant 6$.

For $g = 2, ..., G$, we defined $y^{\text{den}}_{g,i} = 1$ if child i was in grade $g-1$ last year, and 0 otherwise, and defined $y^{\text{num}}_{g,i} = 1$ if child i was in grade $g-1$ last year and is in grade g this year, and 0 otherwise. We defined the totals as $t^{\text{num}}_g = {}_i\, y^{\text{num}}_{g,i}$ and $t^{\text{den}}_g = {}_i\, y^{\text{den}}_{g,i}$, and our estimand was an approximation of our quantity of interest:

$$P(T \geqslant G | T \geqslant 1) \approx \frac{t^{\text{num}}_2 ... t^{\text{num}}_G}{t^{\text{den}}_2 ... t^{\text{den}}_G}.$$

Mortality estimation

Here we describe estimation of the under-five mortality rate. The infant mortality rate was computed analogously. We defined a two-year study period to obtain adequate sample sizes.[12, 4] In 2015, we collected birth and death dates (if the child died) for any child under five that was alive during the study period $[t_{\text{lower}}, t_{\text{upper}}]$, where t_{lower} = January 1, 2013, and t_{upper} = December 31, 2014.

We employed the usual survival setup: T_i = survival time of the ith child, C_i = censoring time of the ith child, and observed data $y_i = \min(T_i, C_i)$, $\nu_i = I(T_i \leqslant C_i)$. Let θ be the parameters of the distribution of T_i and ϕ be the parameters of the distribution of C_i. We derived the posterior distribution of θ by assuming 1) censoring times were independent of survival times, i.e. $C_i \perp T_i$, and 2) distinct parameters, i.e. $p(\theta, \phi) = p(\theta)p(\phi)$:[59]

$$p(\theta | \mathbf{y}, \boldsymbol{\nu}) \propto {}_i \left[p_{T_i}(y_i|\theta)\right]^{\nu_i} \left[S_{T_i}(y_i|\theta)\right]^{1-\nu_i} * p(\theta)$$

where $S_{T_i}(t|\theta) = P(T_i \geqslant t)$ is the survival function. We now describe our model for p_{T_i}, the probability density function of the survival time. Since our quantity of interest is survival until age five (and we do not collect survival data on older children), let any child i that survives beyond age five have $T_i = 5$. Let $\pi = P(T_i = 5)$ be the probability of surviving until age five years.

We defined a distribution which is a mixture of a right-truncated (at five years) Weibull distribution and a probability mass at five years:

$$p_{T_i}(t | \underbrace{\lambda, \kappa, \pi}_{\theta}) = \frac{p_{\text{Weibull}}(t|\kappa, \lambda)}{F_{\text{Weibull}}(5|\kappa, \lambda)} * I(t < 5)(1-\pi) + I(t = 5)\pi$$

We wanted our analysis to be conditional on survival until the start of the study period, t_{lower}. So, we defined A_i as the age of child i at the start of the study period, and as 0 if they were born during the study period: $A_i = \max(t_{\text{lower}} - \text{birth\_date}_i, 0)$. We left-truncated the above mixture distribution to account for the start of the study period:

$$p_{T_i|T_i \geqslant A_i}(t|\theta) = \begin{cases} \frac{p_{T_i}(t|\theta)}{S_{T_i}(A_i|\theta)} & t \geqslant A_i \\ 0 & t < A_i \end{cases}$$

We defined a model for π (the survival probability) by adding levels for mother and household:

$$\pi_i = \text{logit}^{-1}\left(\beta_{m[i]}\right) \text{ for individuals } i$$
$$\beta_m \sim N\left(\alpha_{h[m]}, \sigma^2_{\text{mother}}\right) \text{ for mothers } m$$
$$\alpha_h \sim N\left(\mu, \sigma^2_{\text{household}}\right) \text{ for households } h$$

For inference on the survival probability in a village, we took posterior simulations of $\frac{1}{n} {}^{n}_{i=1} \pi_i$.

Likelihoods can be flat for models defined in restrictive parts of the data space, so we used strong priors on κ and λ based on data from the World Bank.[14] However, inference on the quantities of interest, π_i, was not very sensitive to priors on κ and λ.

6.2 Estimation of quantities of interest

Let $y_{j,t}^{(k)}$ be outcome k in village j and year t, estimated using classical methods. To facilitate comparability, we standardized each outcome (and associated target), subtracting the mean across villages and dividing by the standard deviation across villages, and reoriented it so that larger values indicate higher economic development.

Estimation for a single outcome and country: Let $S_{q,z}^{(k)}$ be the set of villages in country q and treatment group z that include units randomly sampled for outcome k. We estimated the treatment effect on outcome k in country q as

$$\hat{\tau}_q^{(k)} = \mathop{\mathbf{M}}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)} - \mathop{\mathbf{M}}_{j \in S_{q,0}^{(k)}} y_{j,2015}^{(k)}$$

with variance $\frac{1}{|S_{q,1}^{(k)}|} \mathbf{V}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)} + \frac{1}{|S_{q,0}^{(k)}|} \mathbf{V}_{j \in S_{q,0}^{(k)}} y_{j,2015}^{(k)}$, where "M" takes a sample mean over villages, and "V" the sample variance. (Results were not sensitive to whether or not we weighted by village size.) We estimated target attainment as

$$\hat{A}_q^{(k)} = \mathop{\mathbf{M}}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)} - \text{target}_q^{(k)}$$

with variance $\frac{1}{|S_{q,1}^{(k)}|} \mathbf{V}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)}$.

Averaging across outcomes: We created outcome indices for eight categories: (1) poverty, (2) agriculture, (3) nutrition, (4) education, (5) child health, (6) maternal health, (7) HIV and malaria, and (8) water and sanitation. In addition, we created an overall outcome index. A village's outcome index (and associated target) is the average of its standardized outcomes (and targets) within a category. Results were not sensitive to whether we analyzed only villages for which all outcomes were measured or whether indices included only available outcomes. The treatment effect and target attainment were estimated for each outcome index.

Averaging across countries: Treatment effect and target attainment estimates were averaged across countries, weighting all ten countries equally.

6.3 Bayesian multi-outcome model

Let z_j be the indicator of treatment (i.e. that village j is in an MV1). The Bayesian model partially pools information across countries and outcome indices, jointly estimating country-specific treatment effects on the seven indices $m = 1, ..., 7$:[60, 61]

$$\begin{bmatrix} y_{j,2015}^{(1)} \\ \vdots \\ y_{j,2015}^{(7)} \end{bmatrix} \Bigg| \begin{bmatrix} \delta_{q[j]}^{(1)} \\ \vdots \\ \delta_{q[j]}^{(7)} \end{bmatrix}, \begin{bmatrix} \tau_{q[j]}^{(1)} \\ \vdots \\ \tau_{q[j]}^{(7)} \end{bmatrix} \sim N\left(\begin{bmatrix} \delta_{q[j]}^{(1)} + \tau_{q[j]}^{(1)} z_j \\ \vdots \\ \delta_{q[j]}^{(7)} + \tau_{q[j]}^{(7)} z_j \end{bmatrix}, \Sigma_{\text{village}, q[j]} \right) \text{ for villages } j$$

$$\begin{bmatrix} \tau_q^{(m)} \\ \delta_q^{(m)} \end{bmatrix} \Bigg| \begin{bmatrix} \tau \\ \delta \end{bmatrix}, \begin{bmatrix} a_q^{(\tau)} \\ a_q^{(\delta)} \end{bmatrix}, \begin{bmatrix} b_m^{(\tau)} \\ b_m^{(\delta)} \end{bmatrix}, \begin{bmatrix} c^{(\tau)} \\ c^{(\delta)} \end{bmatrix} \sim N\left(\begin{bmatrix} \tau + a_q^{(\tau)} + b_m^{(\tau)} + c^{(\tau)} a_q^{(\tau)} b_m^{(\tau)} \\ \delta + a_q^{(\delta)} + b_m^{(\delta)} + c^{(\delta)} a_q^{(\delta)} b_m^{(\delta)} \end{bmatrix}, \Sigma \right) \text{ for countries } q \text{ and outcome indices } m$$

$$\begin{bmatrix} a_q^{(\tau)} \\ a_q^{(\delta)} \end{bmatrix} \Big| \Gamma_{\text{countries}} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Gamma_{\text{countries}} \right) \text{ for countries } q$$

$$\begin{bmatrix} b_m^{(\tau)} \\ b_m^{(\delta)} \end{bmatrix} \Big| \Gamma_{\text{outcomes}} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Gamma_{\text{outcomes}} \right) \text{ for outcome indices } m$$

We use weakly-informative priors for all remaining parameters, described below. For inference on the treatment effect on each outcome index m in country q, we took posterior simulations of $\tau_q^{(m)}$, averaging to get the effect on the overall index. Models that incorporate uncertainty in village-level estimates produced similar results. Due to geographic homogeneity within each MV1, our matching procedure created narrow matching variable ranges within each country. Given this narrow range, we did not incorporate matching variables into the model. We assessed the model fit via posterior predictive checks, i.e. generating data under the model and comparing to observed data, which show that the model fits well (see Appendix 6.5).[62] We also fit a Bayesian model to one outcome index at a time, see Appendix 6.4.

To fit the above model, we used the following weakly-informative priors:

$$
\begin{aligned}
\Sigma_{\text{village},q} &= \text{diag}(\sigma^{(1)}_{\text{village},q}, ..., \sigma^{(7)}_{\text{village},q})\Omega_{\text{village},q}\text{diag}(\sigma^{(1)}_{\text{village},q}, ..., \sigma^{(7)}_{\text{village},q}) \text{ for countries } q \\
\Sigma &= \text{diag}(\sigma_\tau, \sigma_\delta)\Omega\text{diag}(\sigma_\tau, \sigma_\delta) \\
\Gamma_{\text{countries}} &= \text{diag}(\sigma_{a,\tau}, \sigma_{a,\delta})\Omega_{\text{countries}}\text{diag}(\sigma_{a,\tau}, \sigma_{a,\delta}) \\
\Gamma_{\text{outcomes}} &= \text{diag}(\sigma_{b,\tau}, \sigma_{b,\delta})\Omega_{\text{outcomes}}\text{diag}(\sigma_{b,\tau}, \sigma_{b,\delta}) \\
\sigma_\tau, \sigma_\delta, \sigma_{a,\tau}, \sigma_{a,\delta}, \sigma_{b,\tau}, \sigma_{b,\delta}, \sigma^{(m)}_{\text{village},q} &\sim \text{half-N}(0, 1) \\
\delta, \tau, c^{(\delta)}, c^{(\tau)} &\sim N(0, 1) \\
\Omega_{\text{village},q} &\sim \text{LKJcorr}(3) \text{ for countries } q \\
\Omega_{\text{countries}} &\sim \text{LKJcorr}(3) \\
\Omega_{\text{outcomes}} &\sim \text{LKJcorr}(3).
\end{aligned}
$$

For inference on the treatment effect on outcome index m in country q, we took posterior simulations of $\tau_q^{(m)}$. For inference on the treatment effect on the overall outcome index in country q, we took posterior simulations of $\sum_{m=1}^{7} \tau_q^{(m)} w_m$ where w_m weights by the number of outcomes in each index.

6.4 Bayesian model for a single outcome index

In addition to the joint model defined above, we fit Bayesian models separately to each of the seven outcome indices and the overall outcome index (a total of eight models). Let z_j be the indicator of treatment (i.e. that village j is in an MV1). This single-outcome model partially pools information across countries:[61, 60]

$$
\begin{aligned}
y_{j,2015} \mid \delta_{q[j]}, \tau_{q[j]}, \sigma_{\text{village},q[j]} &\sim N\left(\delta_{q[j]} + \tau_{q[j]} z_j, \sigma^2_{\text{village},q[j]}\right) \text{ for villages } j \\
\begin{bmatrix} \tau_q \\ \delta_q \end{bmatrix} \mid \begin{bmatrix} \tau \\ \delta \end{bmatrix}, \Sigma &\sim N\left(\begin{bmatrix} \tau \\ \delta \end{bmatrix}, \Sigma\right) \text{ for countries } q \\
\sigma_{\text{village},q} \mid \sigma_{\text{village}}, \text{var}_\sigma &\sim \text{LogNormal}\left(\sigma_{\text{village}}, \text{var}_\sigma\right) \text{ for countries } q \\
&\text{weakly informative priors:} \\
\Sigma &= \text{diag}(\sigma_\tau, \sigma_\delta)\Omega\text{diag}(\sigma_\tau, \sigma_\delta) \\
\sigma_\tau, \sigma_\delta, \text{var}_\sigma &\sim \text{half-N}(0, 1) \\
\sigma_{\text{village}}, \delta, \tau &\sim N(0, 1) \\
\Omega &\sim \text{LKJcorr}(3).
\end{aligned}
$$

For inference on the treatment effect in country q, we took posterior simulations of τ_q.

6.5 Posterior predictive checks

We assessed the model fit via posterior predictive checks, i.e. generating data under the model and comparing to observed data.[62] Conditional on the ten countries in our study, we used the model to generate hypothetical data on the overall outcome index for new villages. We plotted these simulated data alongside the observed data in Figure 14, which shows that the model fit well. Results look similar for each individual outcome index.

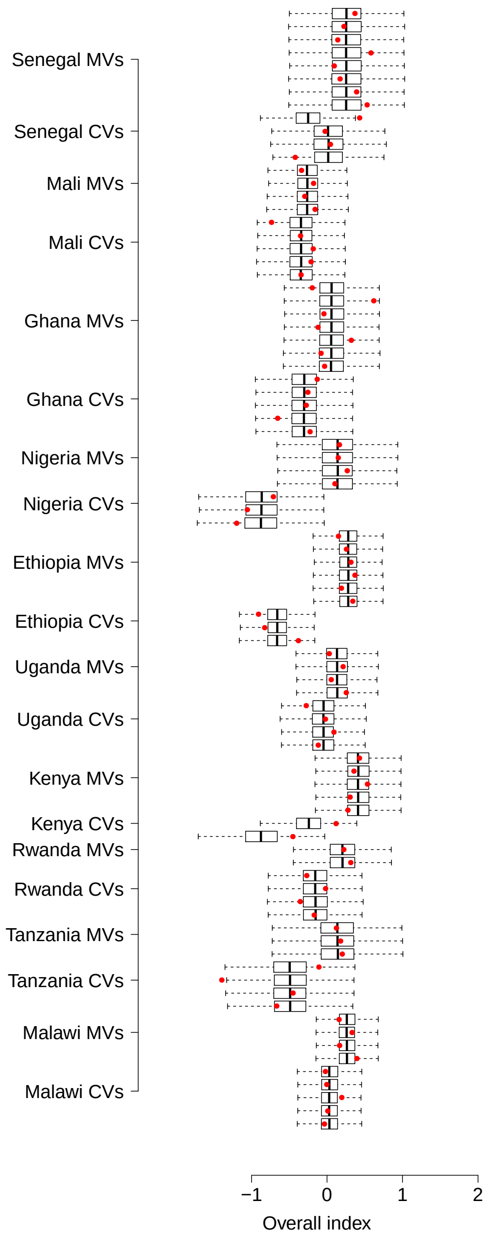


Figure 14: Posterior predictive checks for the multi-outcome, multi-country Bayesian model. Conditional on the ten countries in our study, we used the model to generate hypothetical data on the overall outcome index for new villages across all countries. For each new village, we show a boxplot of the posterior predictive distribution, providing the median, interquartile range (the box), and whiskers (extreme data points that are no more than 1.5 times the interquartile range away from the box). The red dots are overall indices for the observed villages. Observed villages are the random subset of villages for which all outcomes were measured, so that the overall index could be computed.

7 Additional Results

We present additional results to supplement those in the main paper. Figure 15 summarizes the impact evaluation and target attainment classical results, showing the treatment and comparison groups relative to the targets. Figure 16 presents results from a Bayesian model fit to each outcome index separately. Figures 17 through 24 present classical impact results for each country and outcome, and Figures 26 through 33 present classical target attainment results for each country and outcome. See Appendix 9 for corresponding numerical tables.

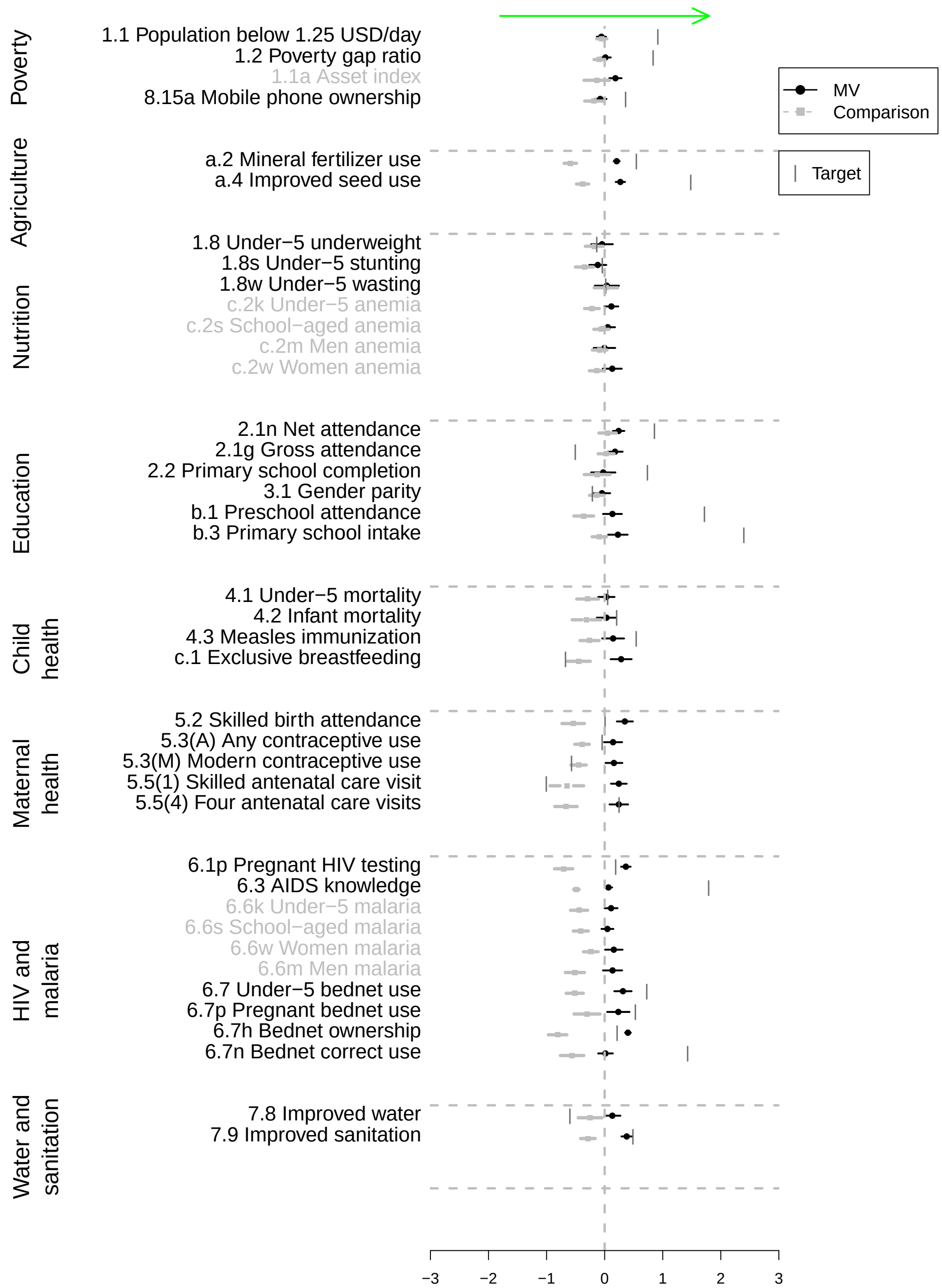
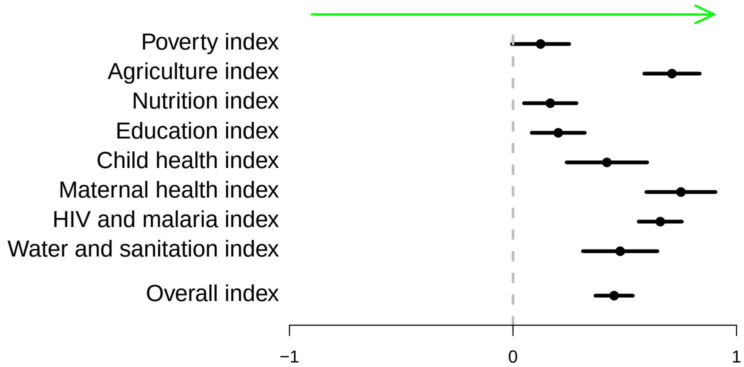


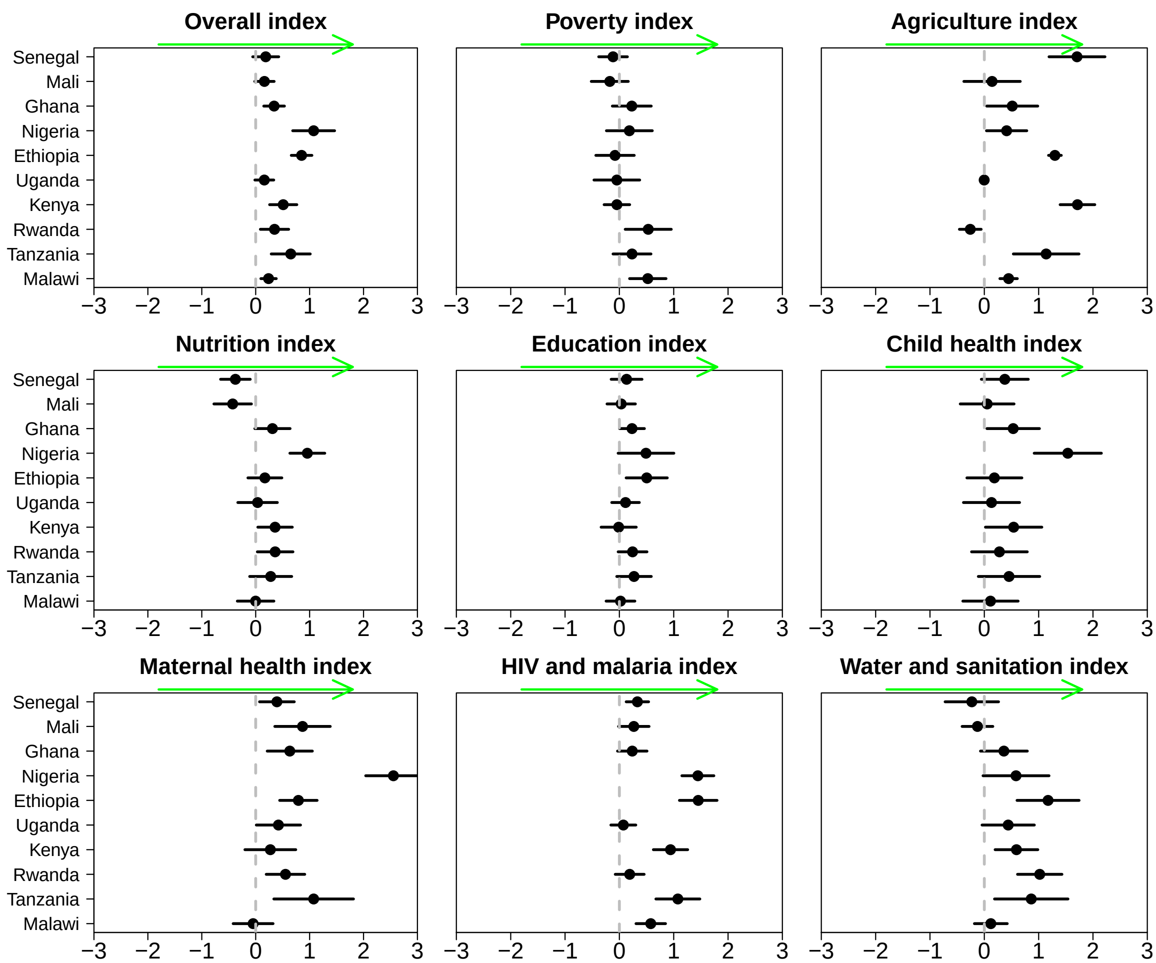
Figure 15: For each outcome, classical estimates and 95% intervals of uncertainty for MVs and comparison villages alongside targets, averaged across countries. As described in the main paper, outcomes were standardized and reoriented. Thus, results are shown as standard deviations from the mean. Outcomes without targets are shown in grey.

7.1 Impact evaluation

Bayesian model for a single outcome index



(a) Averaged across countries



(b) Separately for each country

Figure 16: Estimates of treatment effects and 95% intervals of uncertainty from the Bayesian model defined in Appendix 6.4, fit to each outcome index separately. Results are on the scale of outcome standard deviations.

Classical results

Below are classical estimates of treatment effects with 95% intervals of uncertainty for each outcome and country, presented on the raw scale of the data (rather than in standard deviations).

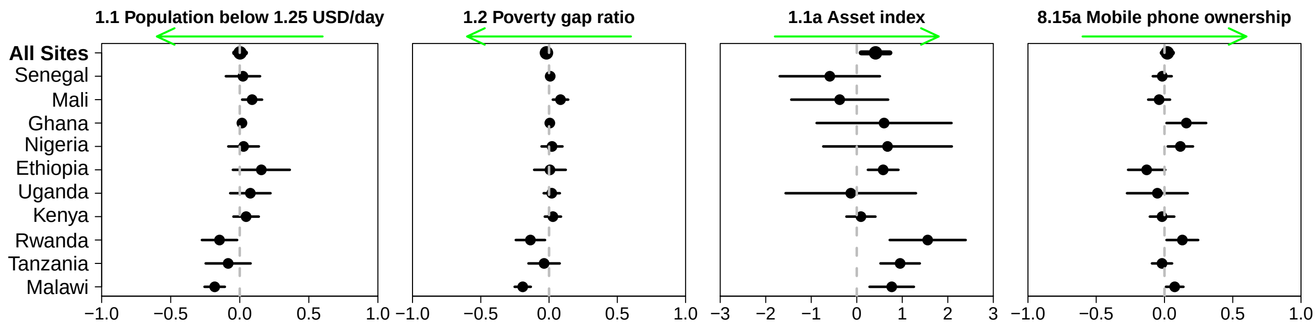


Figure 17: Classical estimates of treatment effects and 95% intervals of uncertainty for the poverty outcomes, presented for each country as well as averaged across countries ("All Sites"). Except for the asset index, all outcomes are proportions and so have possible impacts ranging from -1 to 1. The asset index is the first principal component of a list of assets.[9, 10]

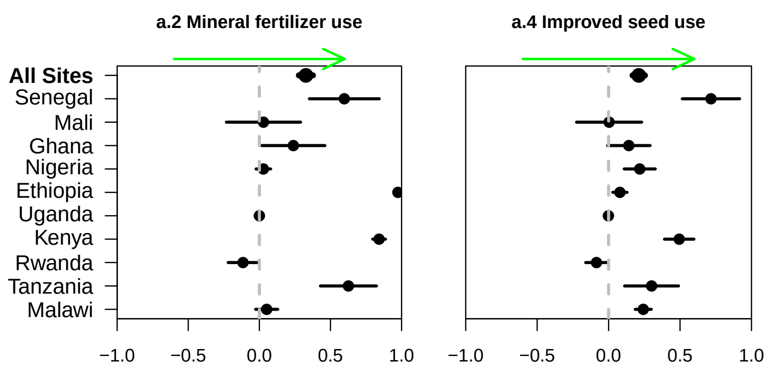


Figure 18: Classical estimates of treatment effects and 95% intervals of uncertainty for the agriculture outcomes, presented for each country as well as averaged across countries ("All Sites"). Both outcomes are proportions and so have possible impacts ranging from -1 to 1.

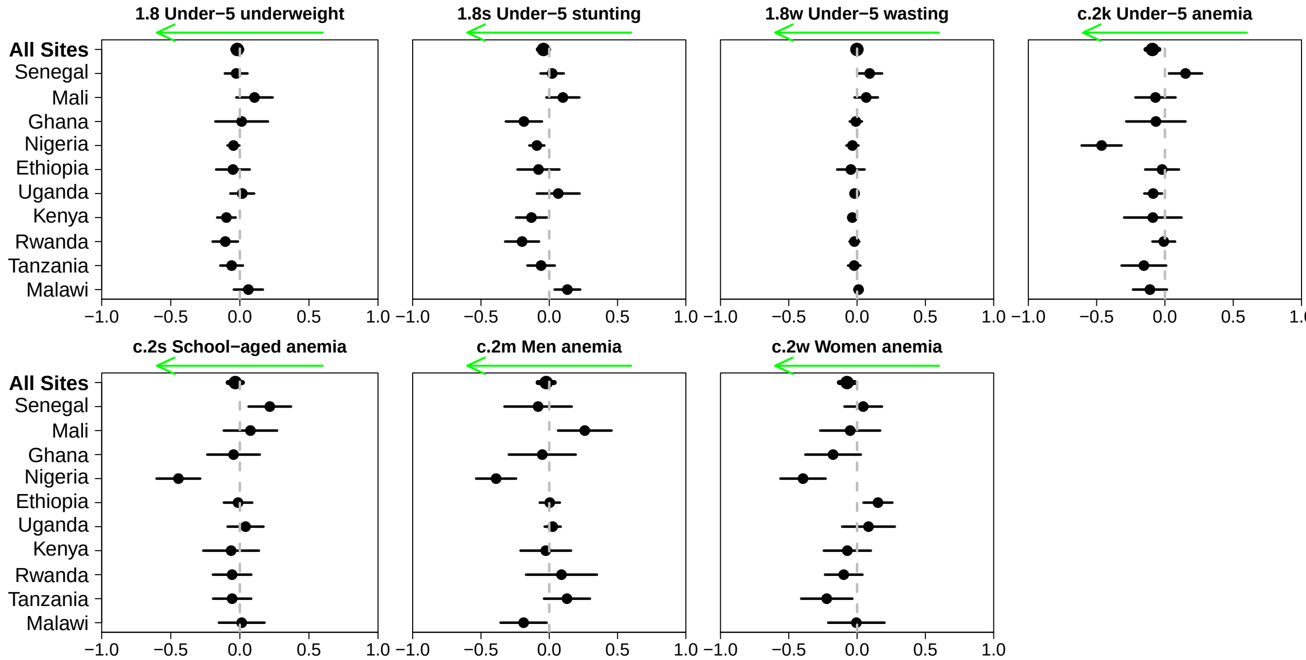


Figure 19: Classical estimates of treatment effects and 95% intervals of uncertainty for the nutrition outcomes, presented for each country as well as averaged across countries ("All Sites"). All outcomes are proportions and so have possible impacts ranging from -1 to 1.

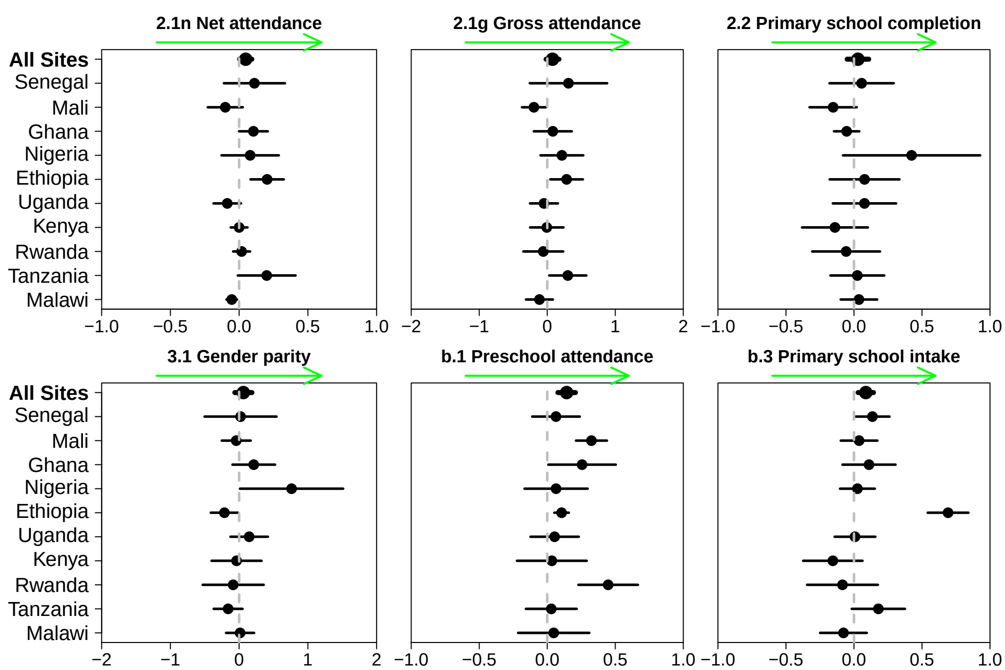


Figure 20: Classical estimates of treatment effects and 95% intervals of uncertainty for the education outcomes, presented for each country as well as averaged across countries ("All Sites"). Except for gross attendance and gender parity, all outcomes are proportions and so have possible impacts ranging from -1 to 1. Gross attendance is the ratio of total attendants in primary school to those of primary school age and so can exceed one. Gender parity is the ratio of girl to boy gross attendance ratios and so can also exceed one. Therefore, both outcomes have possible impacts outside the range from -1 to 1.

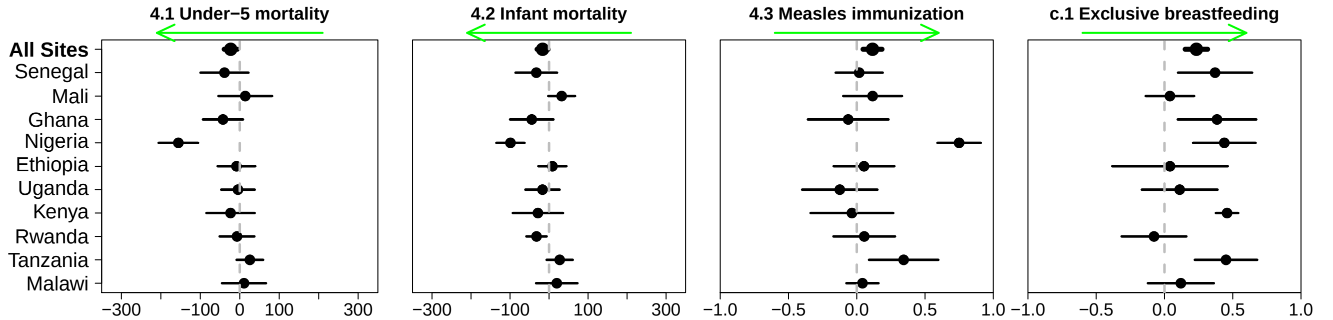


Figure 21: Classical estimates of treatment effects and 95% intervals of uncertainty for the child health outcomes, presented for each country as well as averaged across countries ("All Sites"). Except for the mortality outcomes, all outcomes are proportions and so have possible impacts ranging from -1 to 1. Infant and under-five mortality rates are reported as deaths per 1000 live births.

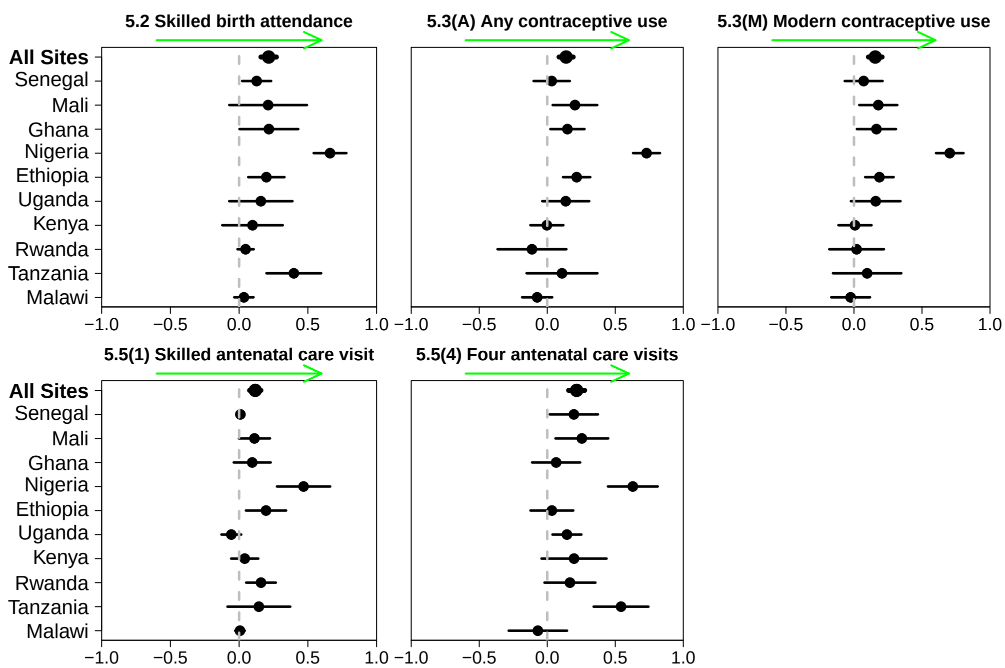


Figure 22: Classical estimates of treatment effects and 95% intervals of uncertainty for the maternal health outcomes, presented for each country as well as averaged across countries ("All Sites"). All outcomes are proportions and so have possible impacts ranging from -1 to 1.

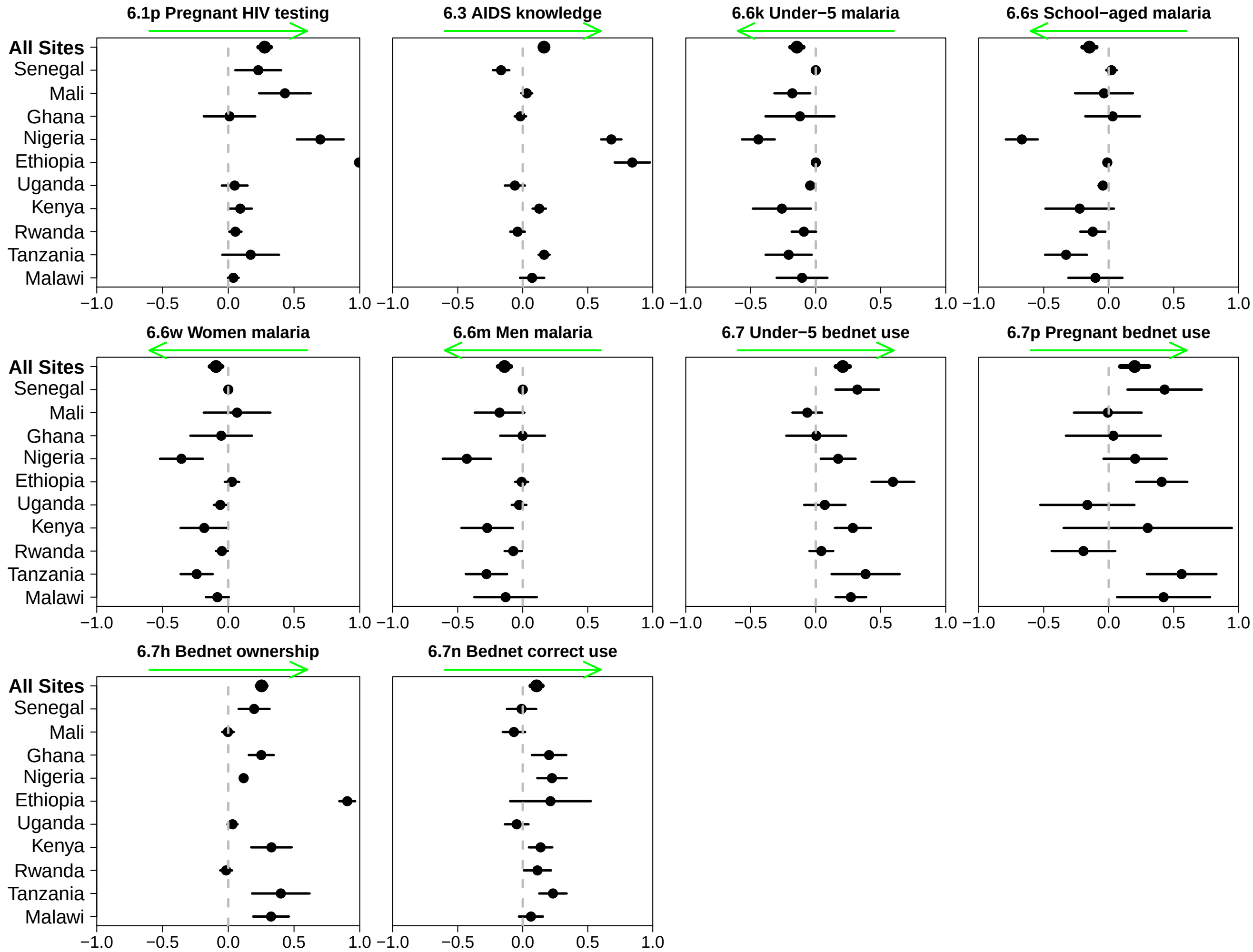


Figure 23: Classical estimates of treatment effects and 95% intervals of uncertainty for the HIV and malaria outcomes, presented for each country as well as averaged across countries ("All Sites"). All outcomes are proportions and so have possible impacts ranging from -1 to 1.

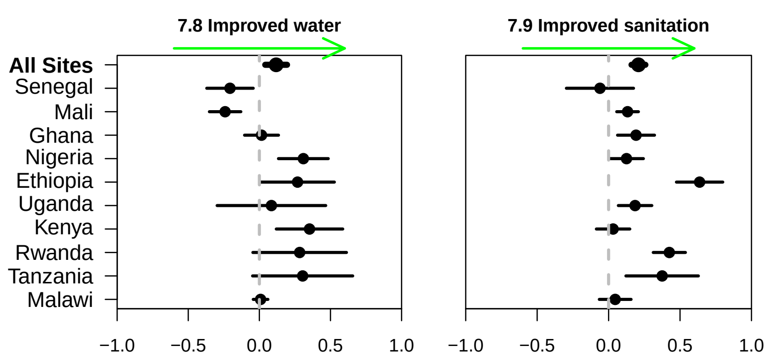


Figure 24: Classical estimates of treatment effects and 95% intervals of uncertainty for the water and sanitation outcomes, presented for each country as well as averaged across countries ("All Sites"). Both outcomes are proportions and so have possible impacts ranging from -1 to 1.

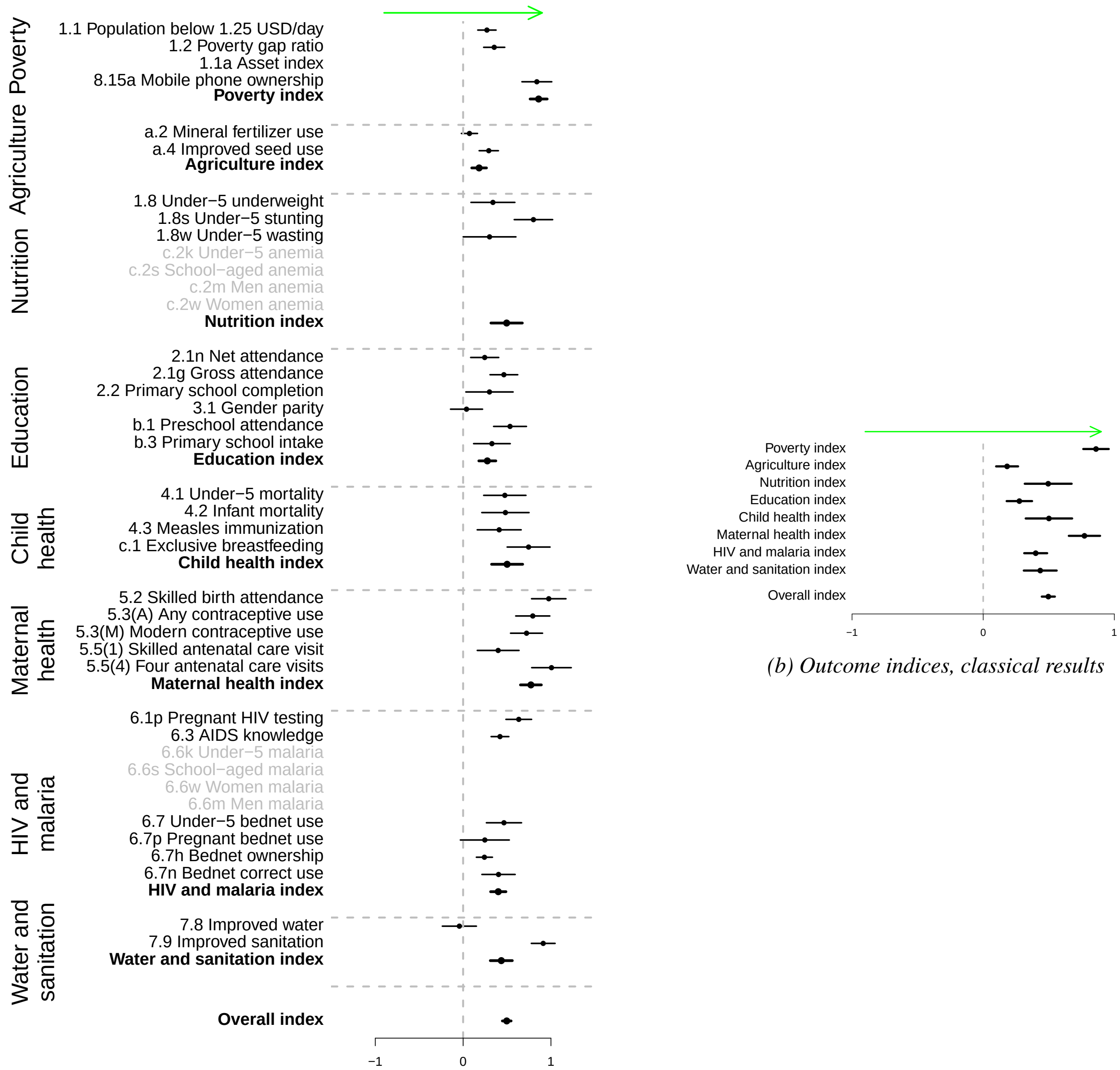
7.2 Time trend estimation and assessment of target attainment

Figure 25 shows the differences between classically-estimated outcomes in 2015 and 2010 cross-sections of the MV1s (with 95% intervals of uncertainty). Analogously to the treatment effect and target attainment estimation, we estimated the change in outcomes over time as

$$\hat{t}_q^{(k)} = \mathop{\mathbf{M}}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)} - \mathop{\mathbf{M}}_{j \in S_{q,1}^{(k)}} y_{j,2010}^{(k)}$$

with variance $\frac{1}{|S_{q,1}^{(k)}|} \mathbf{V}_{j \in S_{q,1}^{(k)}} y_{j,2015}^{(k)} + \frac{1}{|S_{q,1}^{(k)}|} \mathbf{V}_{j \in S_{q,1}^{(k)}} y_{j,2010}^{(k)}$. We calculated the outcome indices over the outcomes that were measured in both years.

Some outcomes were not measured in 2010 and are shown in gray. Averaged across the ten project sites, almost all outcomes improved between 2010 and 2015, with the biggest gains in maternal health outcomes. See Figures 26 to 33 for the results from each outcome and site separately. A limitation of this analysis is that the data collection in 2010 was not identical to the end-line data collection.



(a) Individual outcomes and indices, classical results

(b) Outcome indices, classical results

Figure 25: Estimation of change from 2010 to 2015 and 95% intervals of uncertainty, averaged across countries. As described, outcomes were standardized, reoriented, and subsequently averaged into outcome indices. Thus, results are on the scale of outcome standard deviations. Outcomes not measured in 2010 are shown in grey.

Below are outcome targets and classical estimates of project year five (2010) and ten (2015) outcomes with 95% intervals of uncertainty for each outcome and country, presented on the raw scale of the data (rather than in standard deviations).

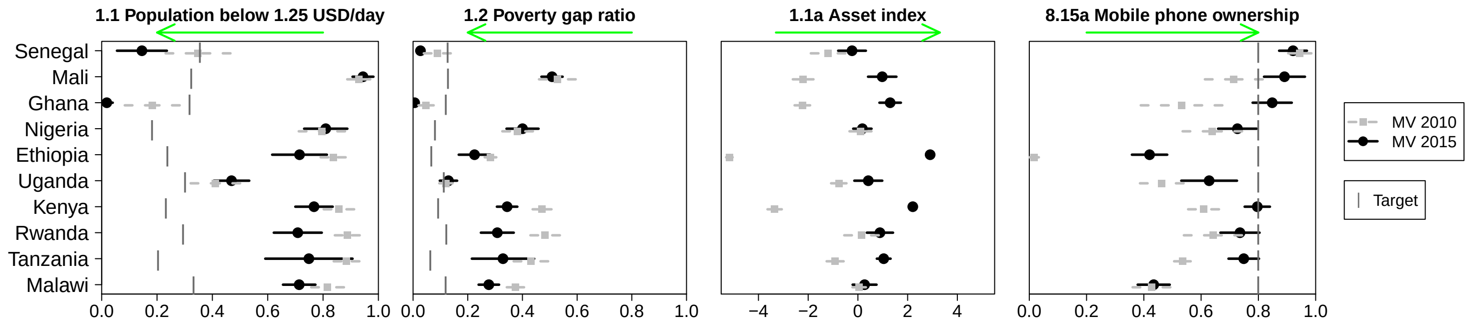


Figure 26: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the poverty outcomes, presented for each country. Except for the asset index, all outcomes are proportions and so range from 0 to 1. The asset index is the first principal component of a list of assets.[9, 10]

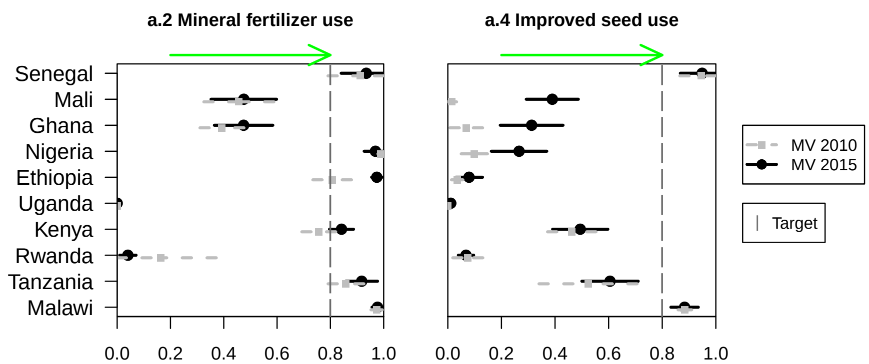


Figure 27: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the agriculture outcomes, presented for each country. Both outcomes are proportions and so range from 0 to 1.

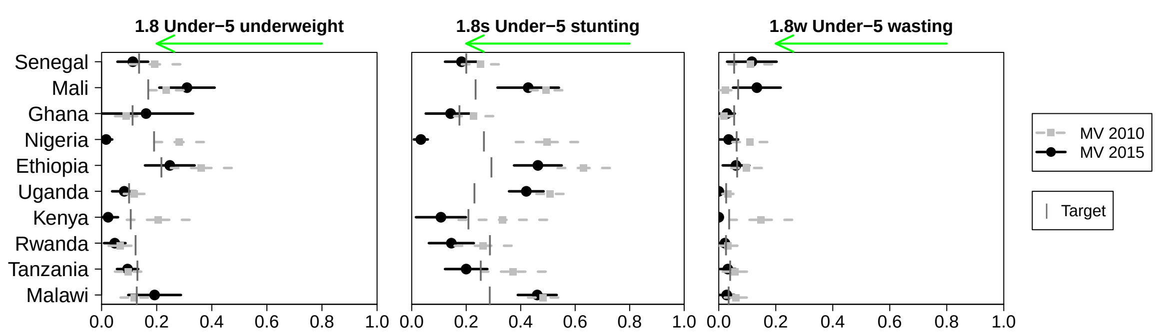


Figure 28: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the nutrition outcomes, presented for each country. All outcomes are proportions and so range from 0 to 1.

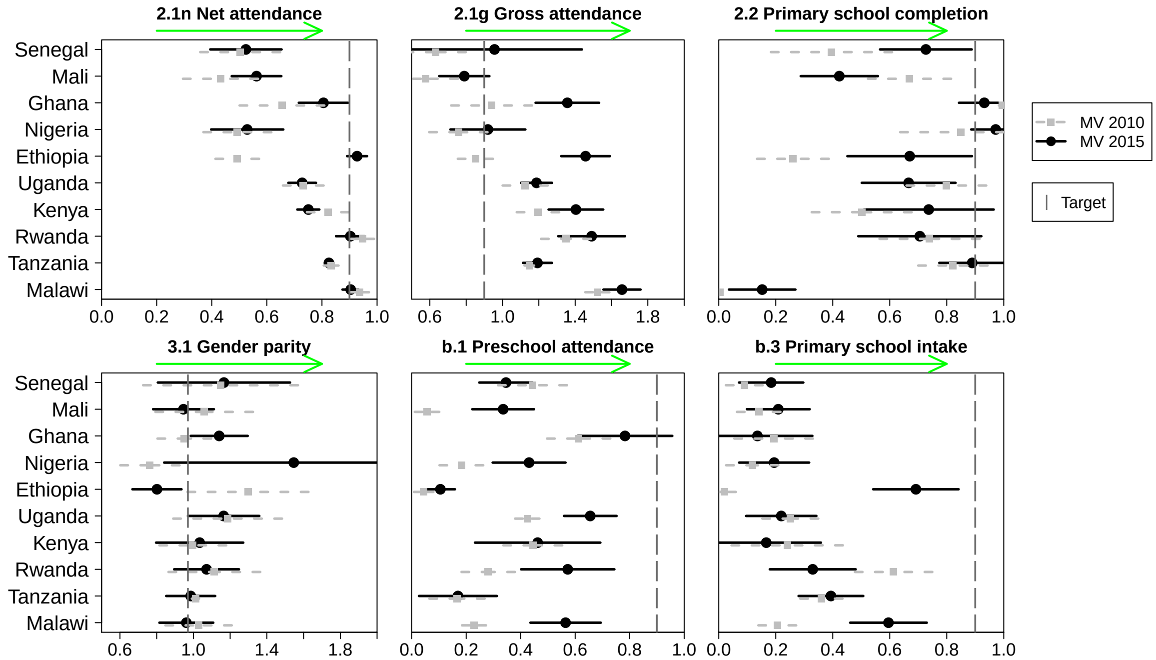


Figure 29: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the education outcomes, presented for each country. Except for gross attendance and gender parity, all outcomes are proportions and so range from 0 to 1. Gross attendance is the ratio of total attendants in primary school to those of primary school age and so can exceed one. Gender parity is the ratio of girl to boy gross attendance ratios and so can also exceed one. Therefore, both outcomes have possible impacts outside the range from -1 to 1.

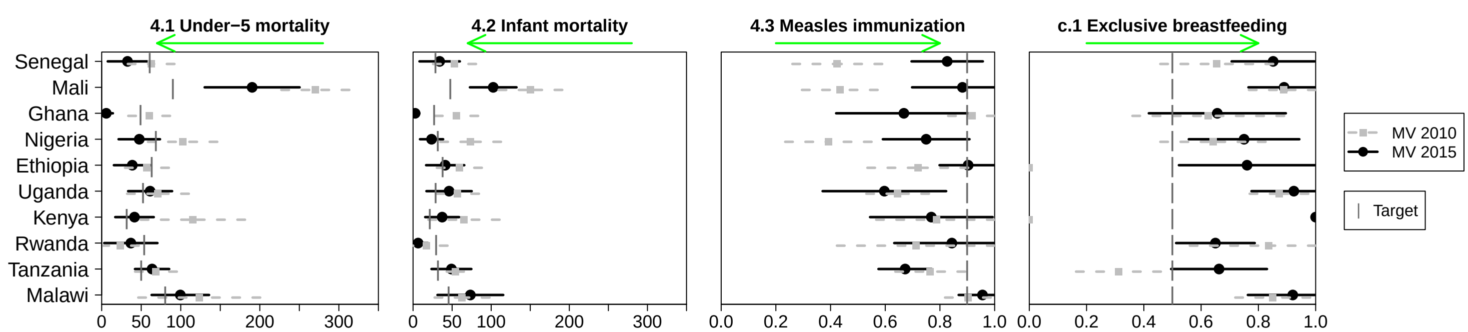


Figure 30: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the child health outcomes, presented for each country. Except for the mortality outcomes, all outcomes are proportions and so range from 0 to 1. Infant and under-five mortality rates are reported as deaths per 1000 live births.

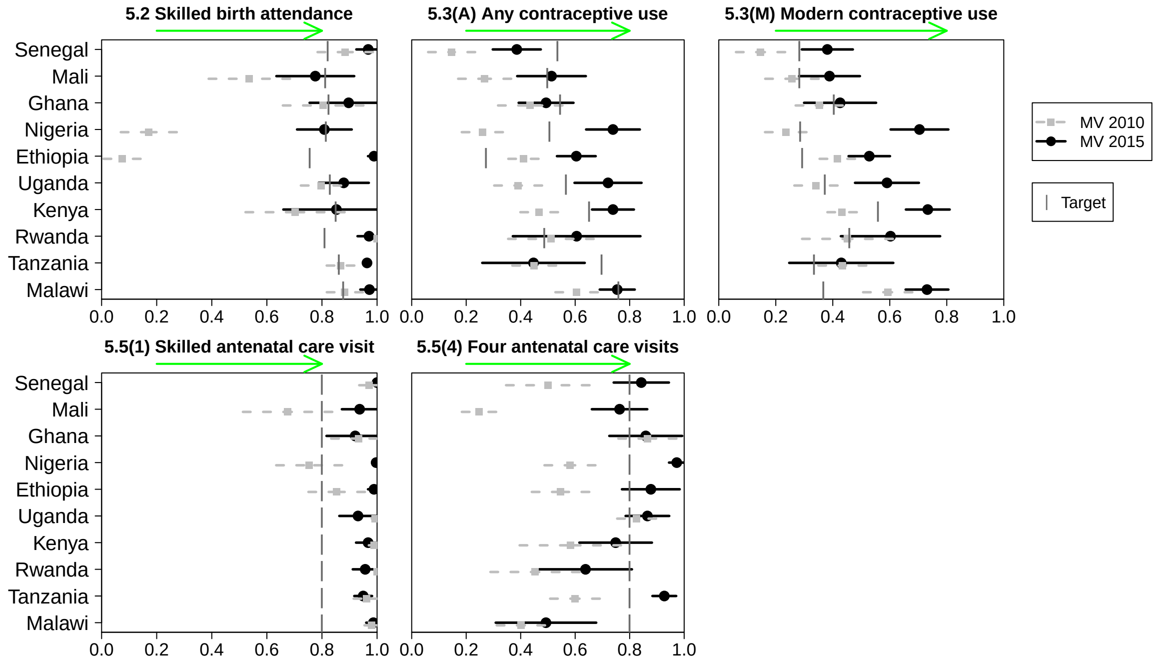


Figure 31: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the maternal health outcomes, presented for each country. All outcomes are proportions and so range from 0 to 1.

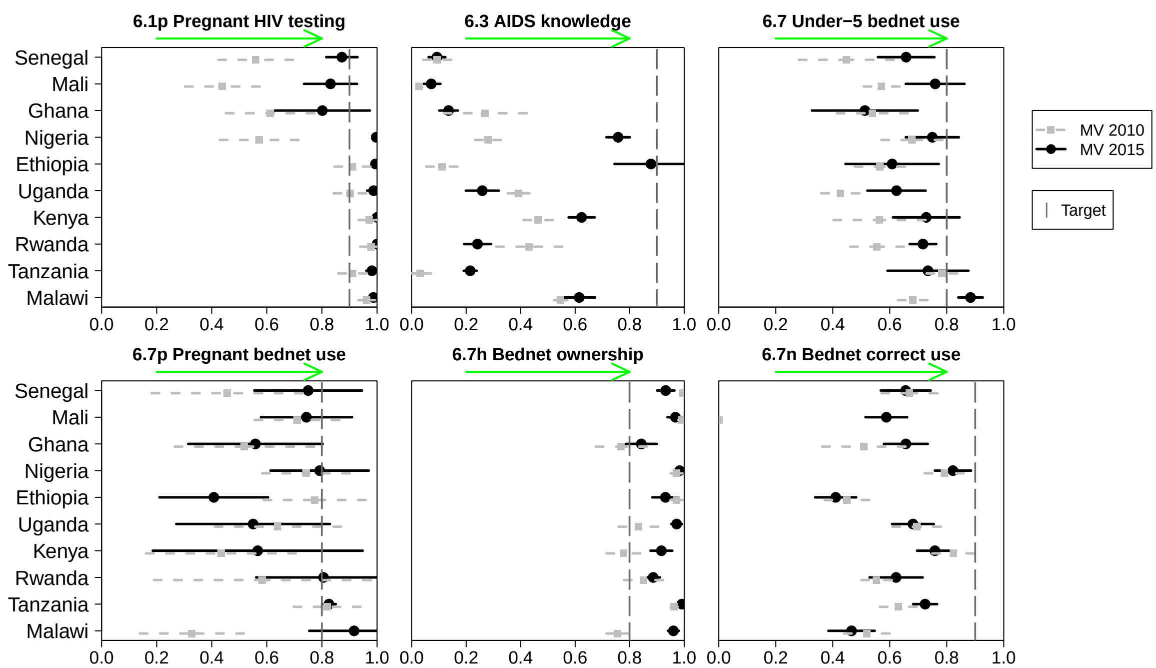


Figure 32: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the HIV and malaria outcomes, presented for each country. All outcomes are proportions and so range from 0 to 1.

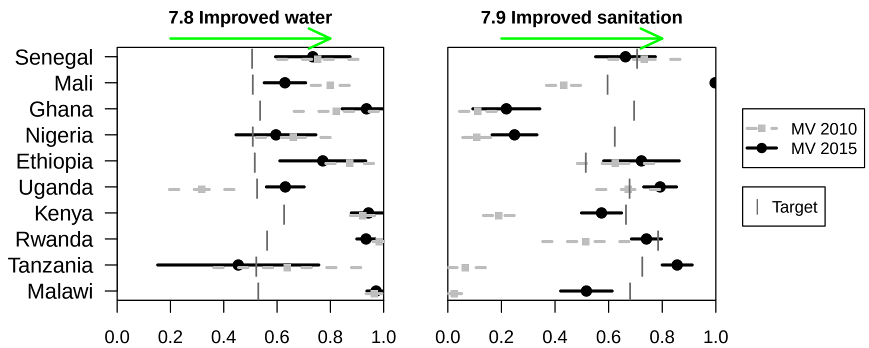


Figure 33: Outcome targets and project year five (2010) and ten (2015) classical estimates with 95% intervals of uncertainty for the water and sanitation outcomes, presented for each country. Both outcomes are proportions and so range from 0 to 1.

8 Assessing Unconfoundedness

Although unconfoundedness cannot be tested directly, there are analyses that can assess its plausibility.[63, 41]. Imbens and Rubin describe an assessment of unconfoundedness using pseudo-outcomes that can be conducted at the design stage (i.e. before outcome data are available). We describe this approach and our results below.

The unconfoundedness assumption is:

$$y(0), y(1) \perp z \mid \mathbf{x} \text{ (unconfoundedness).} \tag{4}$$

A related assumption is subset unconfoundedness, which leaves out the pth pre-treatment variable from the conditioning set:

$$y(0), y(1) \perp z \mid \mathbf{x}^{(-p)} \text{ (subset unconfoundedness).} \tag{5}$$

This assumption cannot be tested for the same reason unconfoundedness cannot be tested: we do not observe $y(1)$ if $z = 0$ and we do not observe $y(0)$ if $z = 1$.[41] Suppose, however, that one of our pre-treatment variables is a good proxy for one of the potential outcomes (e.g. $y(0)$). This variable, $x^{(p)}$, can serve as a pseudo-outcome in a testable version of unconfoundedness:

$$x^{(p)} \perp z \mid \mathbf{x}^{(-p)} \text{ (pseudo-outcome unconfoundedness).} \tag{6}$$

The link between the unconfoundedness assumption (4) and the testable assumption (6) depends on two steps: linking assumptions (4) and (5) and linking assumptions (5) and (6). Both links are based on heuristic arguments that rely on subject-matter knowledge, neither are probabilistic theorems.

While it is theoretically possible that subset unconfoundedness (5) holds but unconfoundedness (4) does not, in practice it is rare if all the $\mathbf{x}$ are pre-treatment variables. Of greater concern is the more plausible scenario that unconfoundedness (4) holds but subset unconfoundedness (5) does not, because conditioning on $x^{(p)}$ is critical.

Subset unconfoundedness (5) and pseudo-outcome unconfoundedness (6) are most closely related when $x^{(p)}$ serves as a good proxy for $y(0)$ or $y(1)$. This is most plausible when $x^{(p)}$ is a lagged version of the outcome.[41] In our analysis, the DHS variables are composites of outcome measures and are therefore some of the best pseudo-outcomes. However, for this same reason they might be critical to condition on, calling into question the subset unconfoundedness assumption.

For pseudo-outcomes, we considered all continuous matching variables, specified at the grid-cell level. We considered nine geographic variables: access to major cities, cation exchange capacity of the soil, percent clay composition of the soil, soil pH, enhanced vegetation index, land surface temperature, elevation, elevation standard deviation (i.e. roughness of terrain), and population density. From the DHS we considered three matching indices: asset wealth, education, and health.

We always included categorical variables (agroecological zone and district or neighboring districts) in $\mathbf{x}^{(-p)}$, and performed exact matching. We then used the continuous variables (except for the pseudo-outcome) in matching procedures described in Appendix 3.2 above. Lastly, we fit a simple hierarchical model,

$$\begin{aligned} x_c^{(p)} &\sim N\left(\delta_{0,q[c]} + \boldsymbol{\delta}\mathbf{x}_c^{(-p)} + \tau z_c, \sigma^2_{\text{grid-cell}}\right) \text{ for grid cells } c \\ \delta_{0,q} &\sim N(\delta_0, \sigma^2_\delta) \text{ for countries } q. \end{aligned} \tag{7}$$

In addition to fitting this model, we also conducted t-tests for the pseudo-outcome between treatment and matched comparison groups.

We performed the above procedure (matching, fitting model 7, and conducting t-tests) for each pseudo-outcome, recording each treatment effect interval of uncertainty. We only had DHS data for both treatment and comparison groups in seven of the ten countries. Therefore, we split our assessment of unconfoundedness into two parts. In one part, we dropped the DHS variables from the pre-treatment variables and pseudo-outcomes and performed the procedure using data from all ten countries. In the other part, we included DHS variables and limited our analysis to data from the seven countries with DHS data.

Without DHS variables, using data from all ten countries, we assessed the nine possible geographic pseudo-outcomes by examining the treatment effect interval of uncertainty from fitting model 7. Only enhanced vegetation index had an interval that did not contain zero. With DHS variables, using data from only seven countries, we assessed all 12 possible pseudo-outcomes by examining the treatment effect interval of uncertainty from fitting model 7. Cation exchange capacity of the soil, elevation, and population density all had intervals that did not contain zero. In addition to fitting model 7, we also conducted t-tests, but none of their results were statistically significant. These four variables were not particularly compelling pseudo-outcomes, therefore we did not abandon attempts at causal inference.

9 Numerical tables of graphs

Tables 5, 6, 7, and 8 include estimates of treatment effects and 95% intervals of uncertainty, averaged across countries. As described, outcomes were standardized, reoriented, and subsequently averaged into outcome indices. Thus, results are on the scale of outcome standard deviations.

Category	Outcome	Estimate (95% uncertainty interval)
Poverty	1.1 Population below 1.25 USD/day	0 (-0.1, 0.1)
	1.2 Poverty gap ratio	0.1 (0, 0.2)
	1.1a Asset index	0.3 (0.1, 0.6)
	8.15a Mobile phone ownership	0.1 (-0.1, 0.3)
	Poverty index	0.1 (0, 0.3)
Agriculture	a.2 Mineral fertilizer use	0.8 (0.7, 0.9)
	a.4 Improved seed use	0.6 (0.5, 0.8)
	Agriculture index	0.7 (0.6, 0.8)
Nutrition	1.8 Under-5 underweight	0.1 (-0.1, 0.4)
	1.8s Under-5 stunting	0.2 (0, 0.4)
	1.8w Under-5 wasting	0 (-0.2, 0.3)
	c.2k Under-5 anemia	0.3 (0.2, 0.5)
	c.2s School-aged anemia	0.1 (-0.1, 0.3)
	c.2m Men anemia	0.1 (-0.1, 0.3)
	c.2w Women anemia	0.3 (0.1, 0.5)
	Nutrition index	0.2 (0.1, 0.3)
Education	2.1n Net attendance	0.2 (0, 0.4)
	2.1g Gross attendance	0.2 (0, 0.3)
	2.2 Primary school completion	0.1 (-0.2, 0.4)
	3.1 Gender parity	0.1 (-0.1, 0.3)
	b.1 Preschool attendance	0.5 (0.3, 0.7)
	b.3 Primary school intake	0.3 (0.1, 0.5)
	Education index	0.2 (0.1, 0.3)
Child health	4.1 Under-5 mortality	0.3 (0.1, 0.6)
	4.2 Infant mortality	0.3 (0, 0.6)
	4.3 Measles immunization	0.4 (0.2, 0.7)
	c.1 Exclusive breastfeeding	0.7 (0.5, 1)
	Child health index	0.4 (0.3, 0.6)
Maternal health	5.2 Skilled birth attendance	0.9 (0.7, 1.1)
	5.3(A) Any contraceptive use	0.5 (0.3, 0.7)
	5.3(M) Modern contraceptive use	0.6 (0.4, 0.8)
	5.5(1) Skilled antenatal care visit	0.9 (0.6, 1.2)
	5.5(4) Four antenatal care visits	0.9 (0.7, 1.2)
	Maternal health index	0.8 (0.6, 0.9)
HIV and malaria	6.1p Pregnant HIV testing	1.1 (0.9, 1.2)
	6.3 AIDS knowledge	0.6 (0.5, 0.6)
	6.6k Under-5 malaria	0.6 (0.4, 0.7)
	6.6s School-aged malaria	0.5 (0.3, 0.6)
	6.6w Women malaria	0.4 (0.2, 0.6)
	6.6m Men malaria	0.6 (0.4, 0.9)
	6.7 Under-5 bednet use	0.8 (0.6, 1)
	6.7p Pregnant bednet use	0.5 (0.2, 0.8)
	6.7h Bednet ownership	1.2 (1, 1.4)
	6.7n Bednet correct use	0.6 (0.3, 0.8)
	HIV and malaria index	0.7 (0.6, 0.8)
Water and sanitation	7.8 Improved water	0.4 (0.1, 0.6)
	7.9 Improved sanitation	0.7 (0.5, 0.8)
	Water and sanitation index	0.5 (0.4, 0.7)
	Overall index	0.5 (0.4, 0.6)

Table 5 (corresponds to Figure 3a in the main paper): individual outcomes and indices, classical results.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	0.1 (0, 0.3)
Agriculture index	0.7 (0.6, 0.8)
Education index	0.2 (0.1, 0.3)
Child health index	0.4 (0.3, 0.6)
Maternal health index	0.8 (0.6, 0.9)
HIV and malaria index	0.7 (0.6, 0.8)
Water and sanitation index	0.5 (0.4, 0.7)
Overall index	0.5 (0.4, 0.6)

Table 6 (corresponds to Figure 3b in the main paper): outcome indices, classical results.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	0.2 (0.1, 0.4)
Agriculture index	0.6 (0.4, 0.7)
Nutrition index	0.2 (0.1, 0.4)
Education index	0.3 (0.2, 0.4)
Child health index	0.5 (0.3, 0.6)
Maternal health index	0.6 (0.4, 0.7)
HIV and malaria index	0.6 (0.5, 0.7)
Water and sanitation index	0.4 (0.3, 0.6)
Overall index	0.4 (0.4, 0.5)

Table 7 (corresponds to Figure 3c in the main paper): outcome indices, Bayesian results.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	0.1 (0, 0.3)
Agriculture index	0.6 (0.5, 0.8)
Nutrition index	0 (-0.1, 0.1)
Education index	0.1 (0, 0.2)
Child health index	0.2 (0, 0.5)
Maternal health index	0.4 (0.3, 0.6)
HIV and malaria index	0.4 (0.3, 0.4)
Water and sanitation index	0.3 (0.1, 0.5)
Overall index	0.3 (0.2, 0.4)

Table 8 (corresponds to Figure 3d in the main paper): outcome indices, classical results, averaged across only the seven countries matched using DHS data.

Tables 9 and 10 include estimates of treatment effects on outcome indices and 95% intervals of uncertainty, separately for each country. Results are on the scale of outcome standard deviations.

	Estimate (95% uncertainty interval)								
Country	Overall	Poverty	Agriculture	Nutrition	Education	Child health	Maternal health	HIV and malaria	Water and sanitation
Senegal	0.1 (-0.2, 0.4)	-0.2 (-0.4, 0.1)	1.8 (1.2, 2.4)	-0.4 (-0.7, -0.2)	0.1 (-0.2, 0.5)	0.4 (-0.2, 0.9)	0.4 (0, 0.7)	0.3 (0.1, 0.5)	-0.4 (-0.9, 0)
Mali	0.1 (0, 0.3)	-0.3 (-0.6, 0)	0 (-0.6, 0.6)	-0.4 (-0.8, -0.1)	0 (-0.2, 0.1)	-0.1 (-0.5, 0.3)	0.9 (0.3, 1.4)	0.2 (0, 0.5)	-0.2 (-0.4, 0)
Ghana	0.3 (0.1, 0.5)	0.3 (-0.2, 0.8)	0.5 (0, 1)	0.4 (0, 0.7)	0.2 (0, 0.5)	0.6 (0, 1.2)	0.6 (0.2, 1)	0.2 (0, 0.5)	0.3 (0, 0.7)
Nigeria	1.2 (0.9, 1.6)	0.2 (-0.3, 0.7)	0.4 (0.2, 0.5)	1.1 (0.9, 1.4)	0.9 (0.1, 1.6)	2 (1.5, 2.4)	2.8 (2.4, 3.3)	1.5 (1.2, 1.8)	0.7 (0.3, 1.1)
Ethiopia	0.9 (0.7, 1.1)	-0.2 (-0.7, 0.3)	1.3 (1.2, 1.4)	0.1 (-0.3, 0.5)	0.7 (0.4, 1)	0.1 (-0.6, 0.8)	0.8 (0.5, 1.1)	1.6 (1.3, 1.8)	1.5 (1, 1.9)
Uganda	0.1 (0, 0.3)	-0.2 (-0.9, 0.5)	0 (0, 0)	-0.1 (-0.3, 0.2)	0.1 (-0.2, 0.3)	0 (-0.5, 0.5)	0.4 (0.1, 0.7)	0 (-0.1, 0.2)	0.4 (-0.3, 1.1)
Kenya	0.5 (0.1, 1)	-0.1 (-0.3, 0.2)	1.8 (1.6, 2)	0.4 (0, 0.8)	-0.2 (-0.5, 0.2)	0.6 (-0.2, 1.4)	0.2 (-0.3, 0.7)	1 (0.8, 1.2)	0.6 (0.1, 1.1)
Rwanda	0.3 (0.2, 0.5)	0.8 (0.3, 1.2)	-0.3 (-0.5, 0)	0.3 (0, 0.7)	0.3 (0, 0.5)	0.3 (-0.2, 0.7)	0.5 (0.3, 0.8)	0.2 (0, 0.3)	1.1 (0.7, 1.6)
Tanzania	0.7 (0.3, 1.2)	0.3 (-0.1, 0.7)	1.2 (0.7, 1.7)	0.3 (0, 0.6)	0.3 (0, 0.7)	0.5 (-0.1, 1)	1.2 (0.2, 2.1)	1.2 (0.7, 1.6)	1.1 (0.3, 1.9)
Malawi	0.2 (0.1, 0.3)	0.6 (0.4, 0.9)	0.4 (0.3, 0.6)	0 (-0.3, 0.3)	-0.1 (-0.2, 0.1)	0 (-0.5, 0.5)	-0.1 (-0.4, 0.2)	0.6 (0.4, 0.8)	0.1 (-0.1, 0.3)

Table 9 (corresponds to Figure 4a in the main paper): classical results.

	Estimate (95% uncertainty interval)								
Country	Overall	Poverty	Agriculture	Nutrition	Education	Child health	Maternal health	HIV and malaria	Water and sanitation
Senegal	0.2 (0, 0.5)	0.1 (-0.3, 0.5)	0.7 (0, 1.5)	-0.4 (-0.8, 0)	0.2 (-0.1, 0.6)	0.3 (-0.2, 0.9)	0.3 (-0.2, 0.7)	0.3 (0, 0.6)	0.2 (-0.2, 0.7)
Mali	0.1 (-0.2, 0.3)	-0.2 (-0.5, 0.2)	0.2 (-0.4, 0.7)	-0.3 (-0.7, 0.1)	0 (-0.3, 0.2)	0 (-0.4, 0.5)	0.8 (0.2, 1.3)	0.4 (0.1, 0.7)	-0.2 (-0.6, 0.1)
Ghana	0.4 (0.1, 0.6)	0.2 (-0.2, 0.7)	0.3 (-0.1, 0.8)	0.3 (0, 0.7)	0.2 (0, 0.5)	0.5 (0, 1)	0.5 (0.1, 1)	0.3 (0, 0.6)	0.3 (-0.1, 0.8)
Nigeria	1 (0.6, 1.3)	0.4 (-0.3, 1.1)	0.3 (0.1, 0.5)	1 (0.6, 1.5)	0.6 (0.1, 1.2)	1 (0.3, 1.8)	1.4 (0.5, 2.3)	1.3 (0.9, 1.8)	1 (0.3, 1.6)
Ethiopia	0.9 (0.7, 1.2)	0.3 (-0.1, 0.7)	1.3 (1.1, 1.4)	0.5 (0.1, 1)	0.8 (0.4, 1.1)	0.6 (0, 1.2)	0.9 (0.5, 1.3)	1.6 (1.3, 1.8)	1 (0.4, 1.6)
Uganda	0.2 (0, 0.4)	0.1 (-0.4, 0.7)	0 (-0.1, 0)	0 (-0.5, 0.4)	0.2 (-0.2, 0.5)	0.4 (0, 0.7)	0.3 (-0.2, 0.8)	0.1 (-0.1, 0.3)	0.3 (-0.3, 0.8)
Kenya	0.7 (0.3, 1)	0.2 (-0.3, 0.7)	1.5 (0.8, 2.2)	0.5 (0, 1)	0.1 (-0.4, 0.6)	0.7 (0, 1.3)	0.5 (-0.2, 1.2)	0.9 (0.5, 1.3)	0.5 (0.1, 0.9)
Rwanda	0.4 (0.1, 0.6)	0.3 (-0.2, 0.9)	0.1 (-0.5, 0.8)	0.2 (-0.2, 0.7)	0.2 (-0.3, 0.7)	0.3 (-0.2, 0.8)	0.5 (0, 0.9)	0.2 (-0.1, 0.6)	0.7 (0, 1.4)
Tanzania	0.6 (0.3, 0.9)	0.3 (-0.1, 0.7)	0.8 (0.2, 1.4)	0.2 (-0.1, 0.6)	0.3 (-0.1, 0.7)	0.5 (0, 1)	0.7 (0, 1.4)	0.9 (0.4, 1.5)	0.7 (0.1, 1.4)
Malawi	0.2 (0.1, 0.4)	0.6 (0.3, 1)	0.4 (0.2, 0.6)	0.1 (-0.3, 0.4)	-0.1 (-0.3, 0.2)	0.1 (-0.4, 0.6)	0 (-0.3, 0.4)	0.5 (0.3, 0.8)	0 (-0.2, 0.2)

Table 10 (corresponds to Figure 4b in the main paper): Bayesian results.

Tables 11 and 12 include the estimated distance from targets and 95% intervals of uncertainty, averaged across countries. As described, outcomes were standardized, reoriented, and subsequently averaged into outcome indices. Thus, results are on the scale of outcome standard deviations.

Category	Outcome	Estimate (95% uncertainty interval)
Poverty	1.1 Population below 1.25 USD/day	-1 (-1.1, -0.9)
	1.2 Poverty gap ratio	-0.8 (-0.9, -0.7)
	8.15a Mobile phone ownership	-0.4 (-0.5, -0.3)
	Poverty index	-0.7 (-0.8, -0.7)
Agriculture	a.2 Mineral fertilizer use	-0.3 (-0.4, -0.3)
	a.4 Improved seed use	-1.2 (-1.3, -1.1)
	Agriculture index	-0.8 (-0.8, -0.7)
Nutrition	1.8 Under-5 underweight	0.1 (-0.1, 0.3)
	1.8s Under-5 stunting	-0.1 (-0.2, 0.1)
	1.8w Under-5 wasting	0 (-0.2, 0.2)
	Nutrition index	0 (-0.1, 0.1)
Education	2.1n Net attendance	-0.6 (-0.7, -0.5)
	2.1g Gross attendance	0.7 (0.6, 0.8)
	2.2 Primary school completion	-0.8 (-1, -0.6)
	3.1 Gender parity	0.2 (0, 0.3)
	b.1 Preschool attendance	-1.6 (-1.7, -1.4)
	b.3 Primary school intake	-2.2 (-2.3, -2)
	Education index	-0.7 (-0.8, -0.6)
Child health	4.1 Under-5 mortality	0 (-0.2, 0.1)
	4.2 Infant mortality	-0.2 (-0.3, 0)
	4.3 Measles immunization	-0.4 (-0.6, -0.2)
	c.1 Exclusive breastfeeding	1 (0.8, 1.1)
	Child health index	0.1 (0, 0.2)
Maternal health	5.2 Skilled birth attendance	0.3 (0.2, 0.5)
	5.3(A) Any contraceptive use	0.2 (0, 0.3)
	5.3(M) Modern contraceptive use	0.7 (0.6, 0.9)
	5.5(1) Skilled antenatal care visit	1.2 (1.1, 1.4)
	5.5(4) Four antenatal care visits	0 (-0.2, 0.2)
	Maternal health index	0.5 (0.4, 0.6)
HIV and malaria	6.1p Pregnant HIV testing	0.2 (0.1, 0.3)
	6.3 AIDS knowledge	-1.7 (-1.8, -1.7)
	6.7 Under-5 bednet use	-0.4 (-0.6, -0.3)
	6.7p Pregnant bednet use	-0.3 (-0.5, -0.1)
	6.7h Bednet ownership	0.2 (0.1, 0.2)
	6.7n Bednet correct use	-1.4 (-1.5, -1.3)
	HIV and malaria index	-0.6 (-0.7, -0.5)
Water and sanitation	7.8 Improved water	0.7 (0.6, 0.9)
	7.9 Improved sanitation	-0.1 (-0.2, 0)
	Water and sanitation index	0.3 (0.2, 0.4)
	Overall index	-0.2 (-0.3, -0.2)

Table 11 (corresponds to Figure 6a in the main paper): individual outcomes and indices, classical results.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	-0.7 (-0.8, -0.7)
Agriculture index	-0.8 (-0.8, -0.7)
Nutrition index	0 (-0.1, 0.1)
Education index	-0.7 (-0.8, -0.6)
Child health index	0.1 (0, 0.2)
Maternal health index	0.5 (0.4, 0.6)
HIV and malaria index	-0.6 (-0.7, -0.5)
Water and sanitation index	0.3 (0.2, 0.4)
Overall index	-0.2 (-0.3, -0.2)

Table 12 (corresponds to Figure 6b in the main paper): outcome indices, classical results.

Tables 13, 14, and 15 include estimated spending in the MV1s in 2005 USD, adjusted using the US Consumer Price Index. We present, per capita: (a) the total spending by site; (b) the total spending by stakeholder, averaged across sites; and (c) the project spending by sector, averaged across sites. Community, donor, and government spending data were collected through 2014; internal project spending data were collected through 2015.

Country	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Senegal	$34	$85	$189	$87	$65	$65	$122	$98	$23	-
Mali	$74	$61	$101	$49	$31	$59	$57	$66	$37	-
Ghana	$83	$156	$179	$352	$235	$226	$259	$155	$135	-
Nigeria	$31	$263	$190	$105	$202	$19	$64	$72	$73	-
Ethiopia	$49	$88	$77	$83	$66	$62	$57	$108	$64	-
Uganda	$91	$103	$83	$77	$106	$62	$91	$79	$65	-
Kenya	$201	$213	$203	$153	$374	$328	$235	$172	$111	-
Rwanda	$131	$184	$171	$151	$159	$96	$109	$199	$135	-
Tanzania	$127	$66	$86	$113	$103	$115	$75	$111	$98	-
Malawi	$132	$183	$146	$178	$113	$101	$114	$134	$95	-

Table 13 (corresponds to Figure 7a in the main paper): total spending, by site.

Stakeholder	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Community	$6	$11	$10	$14	$5	$7	$4	$12	$11	-
Donor	$14	$12	$19	$14	$16	$24	$34	$32	$17	-
Government	$29	$47	$32	$51	$48	$47	$45	$55	$39	-
Project	$47	$70	$81	$56	$76	$34	$34	$21	$17	$18

Table 14 (corresponds to Figure 7b in the main paper): total spending, by stakeholder.

Sector	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Agriculture/Business	$17	$18	$14	$15	$12	$8	$3	$0	$1	$2
Education	$3	$13	$19	$11	$7	$2	$3	$2	$2	$3
Health	$9	$14	$17	$12	$18	$10	$8	$5	$4	$4
Infrastructure	$4	$6	$11	$5	$25	$3	$6	$1	$1	$1
Management	$14	$19	$20	$14	$14	$11	$16	$14	$9	$7

Table 15 (corresponds to Figure 7c in the main paper): project spending, by sector.

	Estimate (95% uncertainty interval)	
Country	MV	Comparison
Senegal	10 (10, 10)	0 (0, 0.1)
Mali	10 (9.9, 10)	0 (0, 0.1)
Ghana	9.8 (9.7, 10)	1.3 (0.5, 2.1)
Nigeria	10 (10, 10)	0 (0, 0)
Ethiopia	9.8 (9.6, 10.1)	0 (0, 0)
Uganda	9.9 (9.8, 10)	0.6 (0.3, 0.9)
Kenya	10 (10, 10)	0.5 (0.1, 0.9)
Rwanda	8.5 (7.9, 9)	1.1 (0.8, 1.4)
Tanzania	9.9 (9.9, 10)	0.4 (0, 0.9)
Malawi	9.7 (9.5, 9.8)	1.1 (0.9, 1.3)

Table 16 (corresponds to Figure 12): classical estimates and 95% intervals of uncertainty for the average number of years lived in the MV site since 2005 for each country, based on household survey of migration patterns.

		Estimate (95% uncertainty interval)		
Category	Outcome	MV	Comparison	Target
Poverty	1.1 Population below 1.25 USD/day	-0.1 (-0.1, 0)	0 (-0.1, 0)	0.9
	1.2 Poverty gap ratio	0 (-0.1, 0.1)	-0.1 (-0.2, 0)	0.8
	1.1a Asset index	0.2 (0.1, 0.3)	-0.1 (-0.4, 0.1)	-
	8.15a Mobile phone ownership	-0.1 (-0.2, 0)	-0.2 (-0.3, 0)	0.4
Agriculture	a.2 Mineral fertilizer use	0.2 (0.2, 0.3)	-0.6 (-0.7, -0.5)	0.5
	a.4 Improved seed use	0.3 (0.2, 0.4)	-0.4 (-0.5, -0.3)	1.5
Nutrition	1.8 Under-5 underweight	0 (-0.2, 0.1)	-0.2 (-0.3, 0)	-0.1
	1.8s Under-5 stunting	-0.1 (-0.3, 0)	-0.3 (-0.5, -0.2)	0
	1.8w Under-5 wasting	0 (-0.2, 0.2)	0 (-0.2, 0.2)	0
	c.2k Under-5 anemia	0.1 (0, 0.2)	-0.2 (-0.3, -0.1)	-
	c.2s School-aged anemia	0.1 (-0.1, 0.2)	-0.1 (-0.2, 0.1)	-
	c.2m Men anemia	0 (-0.2, 0.2)	-0.1 (-0.2, 0)	-
	c.2w Women anemia	0.1 (0, 0.3)	-0.1 (-0.3, 0)	-
Education	2.1n Net attendance	0.2 (0.1, 0.3)	0 (-0.1, 0.2)	0.9
	2.1g Gross attendance	0.2 (0, 0.3)	0 (-0.1, 0.2)	-0.5
	2.2 Primary school completion	0 (-0.2, 0.2)	-0.1 (-0.4, 0.1)	0.7
	3.1 Gender parity	0 (-0.2, 0.1)	-0.1 (-0.3, 0)	-0.2
	b.1 Preschool attendance	0.1 (0, 0.3)	-0.4 (-0.5, -0.2)	1.7
	b.3 Primary school intake	0.2 (0.1, 0.4)	-0.1 (-0.2, 0)	2.4
Child health	4.1 Under-5 mortality	0 (-0.1, 0.2)	-0.3 (-0.5, -0.1)	0.1
	4.2 Infant mortality	0 (-0.1, 0.2)	-0.3 (-0.6, -0.1)	0.2
	4.3 Measles immunization	0.1 (0, 0.3)	-0.3 (-0.4, -0.1)	0.5
	c.1 Exclusive breastfeeding	0.3 (0.1, 0.5)	-0.4 (-0.6, -0.2)	-0.7
Maternal health	5.2 Skilled birth attendance	0.3 (0.2, 0.5)	-0.5 (-0.7, -0.3)	0
	5.3(A) Any contraceptive use	0.1 (0, 0.3)	-0.4 (-0.5, -0.3)	0
	5.3(M) Modern contraceptive use	0.2 (0, 0.3)	-0.4 (-0.6, -0.3)	-0.6
	5.5(1) Skilled antenatal care visit	0.2 (0.1, 0.4)	-0.7 (-0.9, -0.4)	-1
	5.5(4) Four antenatal care visits	0.2 (0.1, 0.4)	-0.7 (-0.9, -0.5)	0.2
HIV and malaria	6.1p Pregnant HIV testing	0.4 (0.3, 0.4)	-0.7 (-0.9, -0.6)	0.2
	6.3 AIDS knowledge	0.1 (0, 0.1)	-0.5 (-0.5, -0.4)	1.8
	6.6k Under-5 malaria	0.1 (0, 0.2)	-0.4 (-0.6, -0.3)	-
	6.6s School-aged malaria	0 (-0.1, 0.2)	-0.4 (-0.5, -0.3)	-
	6.6w Women malaria	0.2 (0, 0.3)	-0.2 (-0.4, -0.1)	-
	6.6m Men malaria	0.1 (0, 0.3)	-0.5 (-0.7, -0.3)	-
	6.7 Under-5 bednet use	0.3 (0.2, 0.5)	-0.5 (-0.7, -0.4)	0.7
	6.7p Pregnant bednet use	0.2 (0, 0.4)	-0.3 (-0.5, -0.1)	0.5
	6.7h Bednet ownership	0.4 (0.4, 0.5)	-0.8 (-1, -0.7)	0.2
	6.7n Bednet correct use	0 (-0.1, 0.1)	-0.6 (-0.8, -0.4)	1.4
Water and sanitation	7.8 Improved water	0.1 (0, 0.3)	-0.2 (-0.5, 0)	-0.6
	7.9 Improved sanitation	0.4 (0.3, 0.5)	-0.3 (-0.4, -0.2)	0.5

Table 17 (corresponds to Figure 15): for each outcome, classical estimates and 95% intervals of uncertainty for MVs and comparison villages alongside targets, averaged across countries. As described in the main paper, outcomes were standardized and reoriented. Thus, results are shown as standard deviations from the mean. Outcomes without targets are shown in grey.

Tables 18 and 19 include estimates of treatment effects and 95% intervals of uncertainty from the Bayesian model defined in Appendix 6.4, fit to each outcome index separately. Results are on the scale of outcome standard deviations.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	0.1 (0, 0.3)
Agriculture index	0.7 (0.6, 0.8)
Nutrition index	0.2 (0.1, 0.3)
Education index	0.2 (0.1, 0.3)
Child health index	0.4 (0.2, 0.6)
Maternal health index	0.8 (0.6, 0.9)
HIV and malaria index	0.7 (0.6, 0.8)
Water and sanitation index	0.5 (0.3, 0.6)
Overall index	0.5 (0.4, 0.5)

Table 18 (corresponds to Figure 16a): Bayesian estimates of treatment effects, averaged across countries.

	Estimate (95% uncertainty interval)								
Country	Overall	Poverty	Agriculture	Nutrition	Education	Child health	Maternal health	HIV and malaria	Water and sanitation
Senegal	0.2 (-0.1, 0.4)	-0.1 (-0.4, 0.1)	1.7 (1.2, 2.2)	-0.4 (-0.7, -0.1)	0.1 (-0.2, 0.4)	0.4 (-0.1, 0.8)	0.4 (0.1, 0.7)	0.3 (0.1, 0.5)	-0.2 (-0.7, 0.3)
Mali	0.2 (0, 0.3)	-0.2 (-0.5, 0.2)	0.1 (-0.4, 0.7)	-0.4 (-0.8, -0.1)	0 (-0.2, 0.3)	0.1 (-0.4, 0.5)	0.9 (0.4, 1.4)	0.3 (0, 0.5)	-0.1 (-0.4, 0.2)
Ghana	0.3 (0.1, 0.5)	0.2 (-0.1, 0.6)	0.5 (0, 1)	0.3 (0, 0.6)	0.2 (0, 0.5)	0.5 (0, 1)	0.6 (0.2, 1.1)	0.2 (0, 0.5)	0.4 (-0.1, 0.8)
Nigeria	1.1 (0.7, 1.5)	0.2 (-0.2, 0.6)	0.4 (0, 0.8)	1 (0.6, 1.3)	0.5 (0, 1)	1.5 (0.9, 2.2)	2.6 (2, 3.1)	1.4 (1.2, 1.7)	0.6 (0, 1.2)
Ethiopia	0.9 (0.7, 1)	-0.1 (-0.4, 0.3)	1.3 (1.2, 1.4)	0.2 (-0.1, 0.5)	0.5 (0.1, 0.9)	0.2 (-0.3, 0.7)	0.8 (0.4, 1.1)	1.5 (1.1, 1.8)	1.2 (0.6, 1.7)
Uganda	0.2 (0, 0.3)	0 (-0.5, 0.4)	0 (0, 0)	0 (-0.3, 0.4)	0.1 (-0.1, 0.4)	0.1 (-0.4, 0.6)	0.4 (0, 0.8)	0.1 (-0.2, 0.3)	0.4 (0, 0.9)
Kenya	0.5 (0.3, 0.8)	0 (-0.3, 0.2)	1.7 (1.4, 2)	0.4 (0, 0.7)	0 (-0.3, 0.3)	0.5 (0, 1.1)	0.3 (-0.2, 0.7)	0.9 (0.6, 1.3)	0.6 (0.2, 1)
Rwanda	0.3 (0.1, 0.6)	0.5 (0.1, 1)	-0.3 (-0.5, -0.1)	0.4 (0, 0.7)	0.2 (0, 0.5)	0.3 (-0.2, 0.8)	0.6 (0.2, 0.9)	0.2 (-0.1, 0.5)	1 (0.6, 1.4)
Tanzania	0.7 (0.3, 1)	0.2 (-0.1, 0.6)	1.1 (0.5, 1.7)	0.3 (-0.1, 0.7)	0.3 (0, 0.6)	0.5 (-0.1, 1)	1.1 (0.3, 1.8)	1.1 (0.7, 1.5)	0.9 (0.2, 1.5)
Malawi	0.2 (0.1, 0.4)	0.5 (0.2, 0.9)	0.4 (0.3, 0.6)	0 (-0.3, 0.3)	0 (-0.2, 0.3)	0.1 (-0.4, 0.6)	0 (-0.4, 0.3)	0.6 (0.3, 0.8)	0.1 (-0.2, 0.4)

Table 19 (corresponds to Figure 16b): Bayesian estimates of treatment effects, separately for each country.

	Estimate (95% uncertainty interval)										
Outcome	All Sites	Senegal	Mali	Ghana	Nigeria	Ethiopia	Uganda	Kenya	Rwanda	Tanzania	Malawi
1.1 Population below 1.25 USD/day	0 (-.04, .04)	.02 (-.1, .15)	.09 (.02, .16)	.02 (-.01, .04)	.03 (-.08, .14)	.16 (-.05, .36)	.08 (-.07, .22)	.05 (-.05, .14)	-.15 (-.27, -.02)	-.08 (-.25, .08)	-.18 (-.26, -.11)
1.2 Poverty gap ratio	-.02 (-.04, .01)	.01 (-.01, .03)	.08 (.03, .14)	.01 (0, .02)	.02 (-.06, .1)	.01 (-.11, .12)	.02 (-.04, .08)	.03 (-.03, .09)	-.14 (-.24, -.03)	-.04 (-.15, .08)	-.19 (-.25, -.13)
1.1a Asset index	.41 (.1, .73)	-.59 (-1.69, .51)	-.37 (-1.44, .69)	.6 (-.88, 2.08)	.68 (-.74, 2.09)	.58 (.24, .92)	-.13 (-1.56, 1.3)	.09 (-.23, .41)	1.56 (.72, 2.39)	.95 (.52, 1.39)	.77 (.28, 1.26)
8.15a Mobile phone ownership	.02 (-.02, .06)	-.02 (-.09, .05)	-.04 (-.12, .04)	.16 (.01, .31)	.12 (.02, .21)	-.13 (-.27, .01)	-.05 (-.27, .17)	-.02 (-.11, .07)	.13 (.01, .25)	-.02 (-.09, .06)	.07 (.01, .14)
a.2 Mineral fertilizer use	.33 (.28, .38)	.6 (.35, .84)	.03 (-.23, .29)	.24 (.02, .46)	.03 (-.02, .08)	.97 (.95, .99)	0 (0, 0)	.84 (.8, .89)	-.12 (-.22, -.01)	.63 (.43, .82)	.05 (-.03, .13)
a.4 Improved seed use	.21 (.17, .25)	.72 (.52, .92)	0 (-.22, .23)	.14 (-.01, .29)	.22 (.11, .33)	.08 (.03, .13)	0 (-.02, .02)	.49 (.39, .6)	-.08 (-.16, -.01)	.3 (.11, .49)	.24 (.19, .3)
1.8 Under-5 underweight	-.02 (-.05, .01)	-.03 (-.11, .06)	.11 (-.03, .24)	.01 (-.18, .2)	-.05 (-.09, 0)	-.05 (-.17, .07)	.02 (-.07, .1)	-.1 (-.16, -.03)	-.11 (-.2, -.01)	-.06 (-.14, .02)	.06 (-.04, .17)
1.8s Under-5 stunting	-.04 (-.08, 0)	.02 (-.06, .11)	.1 (-.02, .22)	-.19 (-.32, -.05)	-.09 (-.15, -.04)	-.08 (-.23, .07)	.07 (-.09, .22)	-.13 (-.24, -.02)	-.2 (-.32, -.07)	-.06 (-.16, .04)	.13 (.04, .23)
1.8w Under-5 wasting	0 (-.02, .02)	.09 (0, .18)	.07 (-.02, .15)	-.01 (-.05, .04)	-.03 (-.08, .01)	-.05 (-.15, .05)	-.02 (-.04, 0)	-.03 (-.07, 0)	-.02 (-.06, .02)	-.02 (-.07, .02)	.01 (-.02, .04)
c.2k Under-5 anemia	-.09 (-.14, -.04)	.15 (.03, .27)	-.07 (-.22, .08)	-.07 (-.28, .15)	-.46 (-.61, -.32)	-.02 (-.14, .1)	-.09 (-.15, -.02)	-.09 (-.3, .12)	-.01 (-.09, .07)	-.15 (-.32, .01)	-.11 (-.23, .01)
c.2s School-aged anemia	-.03 (-.08, .02)	.22 (.06, .37)	.08 (-.12, .27)	-.05 (-.24, .14)	-.44 (-.6, -.29)	-.01 (-.12, .09)	.04 (-.09, .17)	-.06 (-.26, .14)	-.06 (-.2, .08)	-.06 (-.19, .08)	.01 (-.15, .18)
c.2m Men anemia	-.02 (-.08, .04)	-.08 (-.33, .16)	.26 (.06, .46)	-.05 (-.3, .19)	-.39 (-.54, -.24)	0 (-.07, .08)	.03 (-.03, .08)	-.03 (-.21, .16)	.09 (-.17, .35)	.13 (-.04, .3)	-.19 (-.36, -.02)
c.2w Women anemia	-.07 (-.13, -.02)	.05 (-.09, .18)	-.05 (-.27, .17)	-.18 (-.38, .03)	-.4 (-.56, -.23)	.15 (.05, .26)	.08 (-.11, .28)	-.07 (-.24, .1)	-.1 (-.23, .04)	-.22 (-.41, -.03)	-.01 (-.21, .2)
2.1n Net attendance	.05 (0, .09)	.11 (-.11, .33)	-.1 (-.23, .03)	.1 (0, .21)	.08 (-.13, .29)	.2 (.08, .32)	-.09 (-.19, .01)	0 (-.06, .06)	.02 (-.04, .08)	.2 (-.01, .41)	-.05 (-.09, -.02)
2.1g Gross attendance	.08 (-.02, .17)	.31 (-.26, .88)	-.2 (-.37, -.02)	.08 (-.2, .36)	.22 (-.1, .53)	.28 (.04, .52)	-.05 (-.25, .16)	-.01 (-.25, .24)	-.06 (-.35, .23)	.3 (.03, .57)	-.12 (-.31, .08)
2.2 Primary school completion	.03 (-.05, .11)	.06 (-.18, .29)	-.15 (-.33, .02)	-.05 (-.15, .04)	.42 (-.08, .93)	.08 (-.18, .33)	.08 (-.16, .31)	-.14 (-.38, .1)	-.06 (-.31, .19)	.02 (-.17, .22)	.04 (-.1, .17)
3.1 Gender parity	.06 (-.06, .19)	.02 (-.5, .54)	-.04 (-.25, .17)	.21 (-.1, .52)	.76 (.01, 1.51)	-.21 (-.41, -.02)	.15 (-.13, .42)	-.04 (-.4, .33)	-.09 (-.53, .36)	-.16 (-.37, .05)	.01 (-.19, .22)
b.1 Preschool attendance	.14 (.08, .21)	.06 (-.11, .24)	.32 (.21, .44)	.26 (.01, .5)	.06 (-.17, .3)	.11 (.05, .16)	.05 (-.12, .23)	.03 (-.22, .29)	.45 (.23, .67)	.03 (-.16, .22)	.05 (-.21, .31)
b.3 Primary school intake	.09 (.03, .14)	.14 (.01, .26)	.04 (-.1, .17)	.11 (-.08, .31)	.03 (-.1, .15)	.69 (.54, .84)	.01 (-.14, .16)	-.16 (-.37, .06)	-.08 (-.34, .18)	.18 (-.02, .38)	-.08 (-.25, .09)
4.1 Under-5 mortality	-23 (-40, -6)	-39 (-99, 22)	14 (-54, 82)	-43 (-93, 8)	-156 (-205, -106)	-8 (-56, 39)	-5 (-47, 38)	-23 (-84, 38)	-7 (-51, 37)	26 (-8, 59)	11 (-45, 66)
4.2 Infant mortality	-17 (-31, -2)	-33 (-86, 21)	32 (-2, 67)	-44 (-100, 11)	-99 (-135, -63)	9 (-27, 45)	-17 (-61, 27)	-29 (-93, 36)	-32 (-58, -6)	27 (-6, 61)	20 (-33, 73)
4.3 Measles immunization	.12 (.05, .19)	.02 (-.15, .19)	.12 (-.1, .33)	-.06 (-.36, .23)	.75 (.59, .91)	.05 (-.17, .28)	-.12 (-.4, .15)	-.04 (-.34, .27)	.05 (-.17, .28)	.34 (.09, .6)	.04 (-.08, .16)
c.1 Exclusive breastfeeding	.23 (.15, .32)	.37 (.1, .64)	.04 (-.14, .22)	.38 (.1, .67)	.44 (.21, .67)	.04 (-.38, .46)	.11 (-.17, .39)	.46 (.38, .54)	-.08 (-.31, .16)	.45 (.22, .68)	.12 (-.12, .36)
5.2 Skilled birth attendance	.22 (.16, .27)	.13 (.02, .23)	.21 (-.07, .49)	.22 (0, .43)	.66 (.54, .78)	.2 (.07, .33)	.16 (-.07, .39)	.1 (-.12, .32)	.05 (-.01, .11)	.4 (.2, .6)	.03 (-.04, .1)
5.3(A) Any contraceptive use	.14 (.09, .19)	.03 (-.1, .17)	.2 (.04, .37)	.15 (.02, .27)	.73 (.63, .83)	.22 (.12, .32)	.14 (-.04, .31)	0 (-.12, .12)	-.11 (-.37, .14)	.11 (-.15, .37)	-.07 (-.18, .04)
5.3(M) Modern contraceptive use	.16 (.11, .21)	.07 (-.07, .21)	.18 (.04, .32)	.17 (.02, .31)	.7 (.6, .81)	.19 (.08, .29)	.16 (-.02, .34)	.01 (-.11, .13)	.02 (-.18, .22)	.1 (-.15, .35)	-.02 (-.17, .12)
5.5(1) Skilled antenatal care visit	.12 (.08, .16)	.01 (-.01, .02)	.11 (0, .22)	.1 (-.04, .23)	.47 (.28, .66)	.2 (.05, .34)	-.06 (-.13, .02)	.04 (-.06, .14)	.16 (.05, .27)	.14 (-.08, .37)	.01 (-.03, .04)
5.5(4) Four antenatal care visits	.22 (.16, .27)	.2 (.02, .37)	.25 (.06, .45)	.07 (-.11, .24)	.63 (.45, .81)	.03 (-.12, .19)	.14 (.04, .25)	.2 (-.04, .44)	.17 (-.02, .35)	.54 (.34, .74)	-.07 (-.28, .15)
6.1p Pregnant HIV testing	.28 (.23, .32)	.23 (.05, .4)	.43 (.23, .63)	.01 (-.19, .2)	.7 (.52, .88)	.99 (.98, 1)	.05 (-.05, .15)	.09 (0, .18)	.05 (.01, .1)	.17 (-.05, .39)	.04 (0, .08)
6.3 AIDS knowledge	.16 (.14, .19)	-.17 (-.23, -.1)	.03 (-.01, .07)	-.02 (-.06, .03)	.68 (.6, .76)	.84 (.71, .98)	-.06 (-.14, .02)	.13 (.08, .18)	-.04 (-.1, .02)	.16 (.12, .21)	.07 (-.02, .16)
6.6k Under-5 malaria	-.15 (-.19, -.1)	0 (0, 0)	-.18 (-.32, -.04)	-.12 (-.39, .14)	-.44 (-.57, -.32)	0 (-.02, .02)	-.04 (-.07, -.01)	-.26 (-.48, -.04)	-.09 (-.19, 0)	-.21 (-.39, -.03)	-.11 (-.3, .09)
6.6s School-aged malaria	-.15 (-.2, -.1)	.02 (-.02, .06)	-.04 (-.26, .19)	.03 (-.18, .24)	-.67 (-.79, -.55)	-.01 (-.03, .01)	-.04 (-.08, -.01)	-.22 (-.49, .04)	-.12 (-.22, -.03)	-.33 (-.49, -.17)	-.1 (-.31, .11)
6.6w Women malaria	-.09 (-.14, -.05)	0 (0, 0)	.07 (-.19, .32)	-.05 (-.29, .18)	-.36 (-.52, -.19)	.03 (-.03, .08)	-.06 (-.11, -.01)	-.18 (-.36, 0)	-.05 (-.09, 0)	-.24 (-.36, -.12)	-.08 (-.17, 0)
6.6m Men malaria	-.14 (-.19, -.09)	0 (0, 0)	-.18 (-.37, .01)	0 (-.17, .17)	-.43 (-.62, -.25)	-.01 (-.06, .04)	-.03 (-.08, .03)	-.27 (-.47, -.08)	-.07 (-.14, -.01)	-.28 (-.44, -.12)	-.13 (-.37, .11)
6.7 Under-5 bednet use	.21 (.16, .26)	.32 (.15, .49)	-.07 (-.18, .05)	0 (-.23, .23)	.17 (.04, .31)	.59 (.43, .76)	.07 (-.09, .23)	.29 (.15, .42)	.04 (-.05, .13)	.38 (.12, .65)	.27 (.15, .39)
6.7p Pregnant bednet use	.2 (.09, .31)	.43 (.14, .72)	-.01 (-.27, .25)	.04 (-.33, .4)	.2 (-.04, .45)	.41 (.21, .6)	-.16 (-.53, .2)	.3 (-.35, .95)	-.19 (-.44, .05)	.56 (.29, .83)	.42 (.06, .78)
6.7h Bednet ownership	.25 (.22, .29)	.2 (.08, .31)	0 (-.05, .04)	.25 (.16, .35)	.12 (.09, .15)	.9 (.84, .96)	.03 (-.01, .07)	.33 (.17, .48)	-.02 (-.06, .03)	.4 (.18, .62)	.33 (.19, .46)
6.7n Bednet correct use	.11 (.06, .15)	-.01 (-.12, .1)	-.07 (-.15, .02)	.2 (.07, .33)	.22 (.11, .34)	.21 (-.1, .52)	-.05 (-.14, .04)	.14 (.05, .23)	.11 (.01, .22)	.23 (.13, .34)	.06 (-.03, .16)
7.8 Improved water	.12 (.04, .19)	-.21 (-.37, -.04)	-.24 (-.35, -.13)	.02 (-.1, .13)	.31 (.13, .48)	.27 (.01, .53)	.08 (-.3, .47)	.35 (.12, .59)	.28 (-.05, .61)	.3 (-.05, .66)	.01 (-.04, .06)
7.9 Improved sanitation	.21 (.16, .26)	-.06 (-.3, .17)	.13 (.06, .21)	.19 (.06, .32)	.13 (.01, .24)	.64 (.47, .8)	.19 (.07, .3)	.03 (-.09, .15)	.43 (.31, .54)	.37 (.12, .63)	.05 (-.07, .16)

Table 20 (corresponds to Figures 17-24): classical estimates of treatment effects with 95% intervals of uncertainty for each outcome and country, presented on the raw scale of the data (rather than in standard deviations). Most outcomes are proportions and so have possible impacts ranging from -1 to 1. Exceptions are: the asset index, the first principal component of a list of assets;[9, 10] gross attendance, the ratio of total attendants in primary school to those of primary school age; gender parity, the ratio of girl to boy gross attendance ratios; and infant and under-five mortality rates, deaths per 1000 live births.

		Estimate (95% uncertainty interval)									
Outcome	Group	Senegal	Mali	Ghana	Nigeria	Ethiopia	Uganda	Kenya	Rwanda	Tanzania	Malawi
1.1 Population below 1.25 USD/day	MV	.15 (.06, .24)	.94 (.91, .98)	.02 (0, .04)	.81 (.73, .89)	.72 (.62, .81)	.47 (.41, .53)	.77 (.7, .83)	.71 (.62, .8)	.75 (.59, .91)	.71 (.66, .77)
	Comparison	.12 (.04, .21)	.85 (.79, .92)	0 (0, .01)	.78 (.7, .86)	.56 (.38, .74)	.39 (.26, .52)	.72 (.66, .78)	.86 (.76, .95)	.83 (.79, .87)	.9 (.85, .94)
1.2 Poverty gap ratio	MV	.03 (.01, .04)	.51 (.47, .55)	.01 (0, .02)	.4 (.34, .46)	.22 (.17, .28)	.13 (.1, .16)	.34 (.31, .38)	.31 (.25, .37)	.33 (.21, .44)	.28 (.24, .31)
	Comparison	.02 (0, .03)	.42 (.38, .47)	0 (0, 0)	.38 (.33, .43)	.22 (.12, .32)	.11 (.06, .16)	.32 (.27, .36)	.44 (.36, .53)	.37 (.35, .38)	.47 (.42, .52)
1.1a Asset index	MV	-.24 (-.79, .32)	.98 (.41, 1.55)	1.3 (.87, 1.73)	.18 (-.19, .55)	2.91 (2.79, 3.02)	.42 (-.14, .98)	2.21 (2.09, 2.33)	.89 (.36, 1.41)	1.04 (.76, 1.31)	.27 (-.21, .75)
	Comparison	.36 (-.59, 1.31)	1.35 (.45, 2.25)	.7 (-.72, 2.12)	-.5 (-1.86, .87)	2.33 (2.01, 2.64)	.55 (-.77, 1.87)	2.12 (1.82, 2.41)	-.67 (-1.32, -.03)	.09 (-.25, .42)	-.5 (-.58, -.42)
8.15a Mobile phone ownership	MV	.92 (.87, .97)	.89 (.82, .96)	.85 (.78, .92)	.73 (.66, .79)	.42 (.36, .48)	.63 (.53, .73)	.8 (.75, .84)	.74 (.67, .8)	.75 (.7, .8)	.43 (.38, .49)
	Comparison	.94 (.89, .99)	.93 (.89, .97)	.69 (.56, .82)	.61 (.55, .67)	.55 (.43, .67)	.68 (.48, .88)	.81 (.74, .89)	.6 (.51, .7)	.77 (.72, .82)	.36 (.33, .39)
a.2 Mineral fertilizer use	MV	.93 (.84, 1)	.48 (.35, .6)	.47 (.36, .58)	.97 (.93, 1)	.97 (.95, .99)	0 (0, 0)	.84 (.8, .89)	.04 (.01, .07)	.92 (.86, .98)	.98 (.96, 1)
	Comparison	.34 (.11, .56)	.45 (.22, .68)	.24 (.04, .43)	.94 (.91, .97)	0 (0, 0)	0 (0, 0)	0 (0, 0)	.16 (.06, .26)	.29 (.1, .48)	.93 (.85, 1)
a.4 Improved seed use	MV	.95 (.87, 1)	.39 (.29, .49)	.31 (.2, .43)	.27 (.16, .37)	.08 (.03, .13)	.01 (0, .03)	.49 (.39, .6)	.07 (.04, .1)	.61 (.5, .71)	.88 (.83, .93)
	Comparison	.23 (.05, .42)	.39 (.18, .59)	.17 (.08, .26)	.05 (.01, .08)	0 (0, 0)	.01 (0, .03)	0 (0, 0)	.15 (.08, .22)	.3 (.15, .46)	.64 (.62, .66)
1.8 Under-5 underweight	MV	.11 (.06, .17)	.31 (.21, .41)	.16 (0, .33)	.02 (0, .04)	.25 (.16, .34)	.08 (.04, .13)	.02 (0, .06)	.05 (.01, .09)	.09 (.06, .13)	.19 (.1, .29)
	Comparison	.14 (.08, .2)	.2 (.12, .28)	.15 (.06, .24)	.06 (.03, .1)	.3 (.22, .38)	.06 (0, .14)	.12 (.06, .18)	.15 (.07, .24)	.15 (.08, .23)	.13 (.09, .18)
1.8s Under-5 stunting	MV	.18 (.12, .24)	.43 (.31, .54)	.14 (.05, .23)	.03 (.01, .06)	.46 (.38, .55)	.42 (.36, .48)	.11 (.02, .2)	.15 (.06, .23)	.2 (.12, .28)	.46 (.39, .53)
	Comparison	.16 (.1, .22)	.33 (.27, .38)	.33 (.23, .42)	.12 (.08, .17)	.54 (.42, .67)	.35 (.21, .5)	.24 (.17, .3)	.34 (.25, .44)	.26 (.2, .32)	.33 (.27, .39)
1.8w Under-5 wasting	MV	.12 (.03, .2)	.13 (.05, .22)	.03 (0, .06)	.03 (0, .07)	.06 (.01, .11)	0 (0, 0)	0 (0, 0)	.02 (0, .04)	.03 (0, .06)	.03 (0, .05)
	Comparison	.02 (0, .04)	.07 (.05, .08)	.04 (0, .08)	.07 (.04, .1)	.11 (.02, .19)	.02 (0, .04)	.03 (0, .07)	.04 (.01, .07)	.05 (.02, .09)	.02 (0, .03)
c.2k Under-5 anemia	MV	.62 (.52, .72)	.69 (.58, .79)	.46 (.29, .63)	.21 (.11, .31)	.09 (.03, .16)	.04 (.01, .07)	.36 (.21, .51)	.11 (.04, .18)	.28 (.25, .3)	.32 (.23, .41)
	Comparison	.47 (.41, .54)	.75 (.65, .86)	.52 (.39, .66)	.67 (.57, .78)	.11 (.01, .22)	.13 (.07, .19)	.45 (.3, .59)	.12 (.07, .17)	.43 (.27, .59)	.43 (.35, .51)
c.2s School-aged anemia	MV	.74 (.63, .85)	.83 (.72, .93)	.59 (.44, .74)	.39 (.24, .53)	.11 (.02, .19)	.19 (.08, .29)	.27 (.16, .39)	.13 (.04, .22)	.35 (.3, .4)	.51 (.4, .62)
	Comparison	.52 (.42, .63)	.75 (.59, .91)	.63 (.52, .74)	.83 (.76, .9)	.12 (.06, .18)	.14 (.07, .22)	.34 (.17, .5)	.19 (.08, .29)	.41 (.28, .53)	.5 (.38, .62)
c.2m Men anemia	MV	.42 (.23, .62)	.47 (.29, .64)	.28 (.07, .49)	.11 (0, .22)	.04 (0, .1)	.04 (0, .09)	.24 (.09, .4)	.25 (.03, .48)	.49 (.35, .62)	.1 (.01, .2)
	Comparison	.51 (.36, .65)	.21 (.12, .29)	.33 (.21, .45)	.5 (.41, .59)	.04 (0, .08)	.01 (0, .04)	.27 (.18, .36)	.16 (.03, .29)	.36 (.26, .45)	.29 (.15, .43)
c.2w Women anemia	MV	.65 (.53, .76)	.55 (.34, .76)	.34 (.17, .51)	.22 (.08, .37)	.17 (.08, .27)	.23 (.08, .38)	.21 (.09, .34)	.06 (0, .13)	.11 (0, .25)	.29 (.16, .43)
	Comparison	.6 (.52, .68)	.6 (.54, .67)	.52 (.41, .62)	.62 (.54, .7)	.02 (0, .07)	.15 (.03, .27)	.28 (.16, .4)	.16 (.04, .28)	.34 (.2, .47)	.3 (.14, .45)
2.1n Net attendance	MV	.52 (.39, .65)	.56 (.47, .65)	.81 (.72, .9)	.53 (.4, .66)	.93 (.89, .96)	.73 (.68, .78)	.75 (.71, .79)	.9 (.85, .96)	.83 (.81, .84)	.9 (.87, .93)
	Comparison	.41 (.23, .59)	.66 (.57, .75)	.7 (.65, .76)	.45 (.29, .61)	.72 (.61, .84)	.81 (.73, .9)	.75 (.71, .79)	.88 (.85, .92)	.62 (.41, .83)	.96 (.93, .98)
2.1g Gross attendance	MV	.96 (.47, 1.44)	.79 (.65, .93)	1.36 (1.18, 1.53)	.92 (.71, 1.13)	1.46 (1.32, 1.59)	1.19 (1.1, 1.27)	1.4 (1.25, 1.55)	1.49 (1.31, 1.67)	1.19 (1.11, 1.27)	1.66 (1.56, 1.76)
	Comparison	.64 (.34, .95)	.99 (.88, 1.09)	1.28 (1.06, 1.49)	.7 (.47, .94)	1.17 (.97, 1.37)	1.24 (1.05, 1.42)	1.41 (1.22, 1.61)	1.55 (1.32, 1.78)	.89 (.63, 1.15)	1.77 (1.6, 1.94)
2.2 Primary school completion	MV	.73 (.57, .89)	.42 (.29, .56)	.93 (.84, 1)	.97 (.89, 1)	.67 (.45, .89)	.67 (.5, .83)	.74 (.51, .96)	.71 (.49, .92)	.89 (.77, 1)	.15 (.04, .27)
	Comparison	.67 (.51, .83)	.58 (.47, .68)	.99 (.96, 1)	.55 (.07, 1)	.59 (.44, .74)	.59 (.42, .76)	.88 (.77, .99)	.76 (.64, .89)	.86 (.7, 1)	.12 (.05, .18)
3.1 Gender parity	MV	1.17 (.81, 1.53)	.95 (.78, 1.11)	1.14 (.99, 1.3)	1.55 (.84, 2.25)	.8 (.67, .94)	1.16 (.97, 1.36)	1.03 (.8, 1.27)	1.07 (.9, 1.25)	.99 (.85, 1.12)	.96 (.81, 1.11)
	Comparison	1.15 (.77, 1.52)	.99 (.86, 1.12)	.93 (.66, 1.2)	.78 (.53, 1.04)	1.02 (.87, 1.16)	1.02 (.83, 1.21)	1.07 (.8, 1.35)	1.16 (.75, 1.57)	1.15 (.98, 1.31)	.95 (.81, 1.09)
b.1 Preschool attendance	MV	.35 (.25, .44)	.34 (.22, .45)	.78 (.61, .96)	.43 (.3, .56)	.11 (.05, .16)	.66 (.56, .75)	.46 (.23, .69)	.57 (.4, .74)	.17 (.03, .31)	.56 (.44, .69)
	Comparison	.28 (.14, .43)	.01 (0, .02)	.53 (.35, .71)	.37 (.18, .55)	0 (0, 0)	.6 (.46, .75)	.43 (.32, .54)	.13 (0, .26)	.14 (.02, .26)	.52 (.29, .74)
b.3 Primary school intake	MV	.18 (.07, .3)	.21 (.1, .32)	.14 (0, .33)	.19 (.07, .32)	.69 (.54, .84)	.22 (.1, .34)	.17 (0, .36)	.33 (.18, .48)	.39 (.28, .51)	.6 (.46, .73)
	Comparison	.05 (0, .1)	.17 (.09, .25)	.03 (0, .06)	.17 (.13, .2)	0 (0, 0)	.21 (.13, .3)	.32 (.22, .43)	.41 (.2, .63)	.21 (.05, .37)	.67 (.57, .78)
4.1 Under-5 mortality	MV	33 (8, 58)	190 (130, 250)	6 (-3, 14)	48 (21, 74)	39 (15, 62)	61 (33, 89)	42 (17, 66)	37 (3, 71)	64 (42, 85)	99 (63, 136)
	Comparison	71 (16, 127)	176 (144, 208)	48 (-2, 98)	203 (161, 246)	47 (6, 89)	66 (34, 98)	65 (9, 121)	44 (16, 72)	38 (12, 64)	89 (46, 131)
4.2 Infant mortality	MV	34 (8, 60)	103 (73, 132)	3 (-3, 8)	24 (9, 39)	41 (17, 66)	46 (17, 75)	37 (16, 59)	7 (-6, 19)	49 (24, 75)	73 (31, 115)
	Comparison	67 (20, 113)	70 (53, 87)	47 (-8, 102)	123 (90, 155)	33 (6, 59)	63 (30, 96)	66 (5, 126)	39 (16, 61)	22 (0, 43)	54 (21, 86)
4.3 Measles immunization	MV	.83 (.7, .96)	.88 (.7, 1)	.67 (.42, .92)	.75 (.59, .91)	.9 (.8, 1)	.6 (.37, .82)	.77 (.54, .99)	.84 (.63, 1)	.67 (.58, .77)	.96 (.87, 1)
	Comparison	.81 (.71, .91)	.77 (.66, .88)	.73 (.57, .89)	0 (0, 0)	.85 (.65, 1)	.72 (.58, .86)	.8 (.6, 1)	.79 (.66, .92)	.33 (.08, .58)	.91 (.84, .99)
c.1 Exclusive breastfeeding	MV	.85 (.71, 1)	.89 (.77, 1)	.66 (.42, .9)	.75 (.56, .94)	.76 (.52, 1)	.92 (.78, 1)	1 (1, 1)	.65 (.51, .79)	.66 (.5, .83)	.92 (.76, 1)
	Comparison	.48 (.25, .71)	.85 (.72, .98)	.27 (.11, .43)	.31 (.19, .43)	.72 (.37, 1)	.81 (.58, 1)	.54 (.46, .62)	.73 (.53, .92)	.21 (.06, .37)	.8 (.62, .98)
5.2 Skilled birth attendance	MV	.97 (.92, 1)	.78 (.63, .92)	.9 (.76, 1)	.81 (.71, .91)	.99 (.97, 1)	.88 (.79, .97)	.85 (.66, 1)	.97 (.93, 1)	.96 (.96, .97)	.97 (.94, 1)
	Comparison	.84 (.74, .94)	.56 (.32, .81)	.68 (.52, .84)	.15 (.08, .21)	.79 (.66, .92)	.72 (.51, .93)	.76 (.65, .86)	.92 (.88, .97)	.57 (.37, .76)	.94 (.88, 1)
5.3(A) Any contraceptive use	MV	.39 (.3, .47)	.51 (.39, .64)	.49 (.39, .59)	.74 (.64, .84)	.6 (.53, .68)	.72 (.6, .84)	.74 (.66, .82)	.61 (.37, .84)	.45 (.26, .64)	.75 (.69, .82)
	Comparison	.35 (.25, .45)	.31 (.2, .41)	.34 (.27, .42)	.01 (0, .02)	.39 (.32, .46)	.58 (.46, .71)	.74 (.65, .84)	.72 (.62, .81)	.34 (.16, .52)	.83 (.74, .92)
5.3(M) Modern contraceptive use	MV	.38 (.29, .47)	.39 (.28, .5)	.43 (.3, .55)	.7 (.6, .81)	.53 (.46, .6)	.59 (.48, .7)	.73 (.66, .81)	.6 (.43, .78)	.43 (.25, .61)	.73 (.66, .81)
	Comparison	.31 (.2, .42)	.21 (.12, .3)	.26 (.19, .33)	0 (0, 0)	.34 (.27, .42)	.43 (.29, .57)	.73 (.63, .82)	.58 (.48, .68)	.33 (.16, .51)	.76 (.63, .88)
5.5(1) Skilled antenatal care visit	MV	1 (1, 1)	.94 (.87, 1)	.92 (.82, 1)	1 (.99, 1)	.99 (.97, 1)	.93 (.86, 1)	.97 (.92, 1)	.96 (.91, 1)	.95 (.92, .98)	.99 (.96, 1)
	Comparison	.99 (.98, 1)	.83 (.73, .92)	.83 (.74, .91)	.53 (.33, .72)	.79 (.65, .94)	.99 (.96, 1)	.93 (.84, 1)	.8 (.7, .9)	.81 (.58, 1)	.98 (.96, 1)
5.5(4) Four antenatal care visits	MV	.84 (.74, .94)	.76 (.66, .86)	.86 (.73, .99)	.97 (.94, 1)	.88 (.77, .98)	.87 (.79, .95)	.75 (.62, .88)	.64 (.47, .81)	.93 (.88, .97)	.49 (.31, .68)
	Comparison	.65 (.5, .79)	.51 (.34, .67)	.79 (.68, .91)	.34 (.16, .52)	.84 (.73, .96)	.72 (.65, .79)	.55 (.36, .75)	.47 (.39, .55)	.38 (.19, .58)	.56 (.45, .67)
6.1p Pregnant HIV testing	MV	.87 (.82, .93)	.83 (.73, .93)	.8 (.63, .97)	1 (.99, 1)	.99 (.98, 1)	.99 (.96, 1)	1 (1, 1)	1 (1, 1)	.98 (.96, 1)	.99 (.96, 1)
	Comparison	.64 (.48, .81)	.4 (.23, .57)	.79 (.7, .88)	.3 (.12, .47)	0 (0, 0)	.94 (.84, 1)	.91 (.82, 1)	.95 (.9, .99)	.81 (.6, 1)	.95 (.92, .98)
6.3 AIDS knowledge	MV	.09 (.06, .12)	.07 (.04, .11)	.14 (.1, .17)	.76 (.71, .8)	.88 (.74, 1)	.26 (.2, .32)	.62 (.58, .67)	.24 (.19, .29)	.21 (.19, .24)	.61 (.56, .67)
	Comparison	.26 (.2, .31)	.04 (.02, .07)	.15 (.13, .18)	.08 (.01, .14)	.04 (.01, .06)	.32 (.27, .37)	.5 (.48, .52)	.28 (.25, .31)	.05 (.02, .09)	.54 (.47, .62)
6.6k Under-5 malaria	MV	0 (0, 0)	.41 (.35, .47)	.37 (.19, .55)	.15 (.08, .22)	.01 (0, .02)	0 (0, 0)	.43 (.26, .59)	0 (0, 0)	.18 (.11, .26)	.38 (.28, .49)
	Comparison	0 (0, 0)	.59 (.46, .72)	.49 (.3, .68)	.59 (.49, .7)	.01 (0, .02)	.04 (.01, .07)	.69 (.53, .84)	.09 (0, .19)	.39 (.23, .55)	.49 (.32, .65)
6.6s School-aged malaria	MV	.02 (0, .06)	.76 (.58, .93)	.59 (.43, .74)	.02 (0, .05)	0 (0, 0)	0 (0, 0)	.46 (.34, .58)	0 (0, 0)	.16 (.13, .2)	.46 (.31, .61)
	Comparison	0 (0, 0)	.79 (.66, .93)	.56 (.41, .7)	.69 (.57, .81)	.01 (0, .03)	.04 (.01, .08)	.68 (.45, .92)	.12 (.03, .22)	.49 (.33, .65)	.56 (.41, .71)
6.6w Women malaria	MV	0 (0, 0)	.52 (.31, .74)	.23 (.04, .43)	.15 (.04, .25)	.03 (0, .08)	0 (0, 0)	.26 (.12, .39)	0 (0, 0)	.01 (0, .03)	.12 (.04, .2)
	Comparison	0 (0, 0)	.46 (.32, .59)	.29 (.15, .42)	.5 (.38, .63)	0 (0, 0)	.06 (.01, .11)	.44 (.31, .56)	.05 (0, .09)	.25 (.13, .37)	.2 (.16, .24)
6.6m Men malaria	MV	0 (0, 0)	.41 (.28, .55)	.21 (.06, .36)	.04 (0, .12)	.02 (0, .05)	0 (0, 0)	.13 (.01, .25)	0 (0, 0)	.08 (0, .16)	.2 (0, .41)
	Comparison	0 (0, 0)	.59 (.46, .73)	.21 (.13, .29)	.47 (.31, .64)	.03 (0, .06)	.03 (0, .08)	.4 (.26, .55)	.07 (.01, .14)	.36 (.22, .5)	.34 (.21, .46)
6.7 Under-5 bednet use	MV	.66 (.56, .76)	.76 (.66, .86)	.51 (.33, .7)	.75 (.66, .84)	.61 (.44, .77)	.62 (.52, .73)	.73 (.61, .85)	.72 (.67, .76)	.73 (.59, .88)	.88 (.84, .93)
	Comparison	.34 (.2, .47)	.82 (.78, .87)	.51 (.37, .64)	.58 (.48, .67)	.01 (0, .03)	.55 (.43, .68)	.44 (.37, .52)	.67 (.6, .75)	.35 (.13, .57)	.61 (.5, .72)
6.7p Pregnant bednet use	MV	.75 (.55, .95)	.74 (.58, .91)	.56 (.31, .8)	.79 (.61, .97)	.41 (.21, .6)	.55 (.27, .83)	.57 (.19, .95)	.81 (.56, 1)	.83 (.8, .85)	.92 (.75, 1)
	Comparison	.32 (.11, .53)	.75 (.55, .95)	.52 (.25, .79)	.59 (.42, .75)	0 (0, 0)	.71 (.48, .94)	.27 (0, .79)	1 (1, 1)	.26 (0, .53)	.49 (.18, .81)
6.7h Bednet ownership	MV	.93 (.9, .96)	.97 (.94, 1)	.84 (.79, .9)	.98 (.97, 1)	.93 (.88, .98)	.97 (.95, .99)	.92 (.88, .96)	.89 (.86, .91)	.99 (.98, 1)	.96 (.94, .98)
	Comparison	.74 (.62, .85)	.97 (.94, 1)	.59 (.52, .67)	.87 (.84, .89)	.03 (0, .06)	.94 (.91, .97)	.59 (.44, .74)	.9 (.87, .94)	.59 (.37, .81)	.64 (.5, .77)
6.7n Bednet correct use	MV	.66 (.57, .74)	.59 (.51, .66)	.66 (.58, .73)	.82 (.76, .89)	.41 (.34, .48)	.68 (.61, .75)	.76 (.7, .82)	.62 (.53, .72)	.72 (.68, .77)	.47 (.38, .55)
	Comparison	.66 (.59, .74)	.66 (.61, .7)	.45 (.35, .56)	.6 (.5, .69)	.2 (0, .5)	.73 (.67, .78)	.62 (.56, .69)	.51 (.46, .56)	.49 (.4, .59)	.4 (.36, .45)
7.8 Improved water	MV	.73 (.59, .87)	.63 (.55, .71)	.94 (.84, 1)	.6 (.45, .75)	.77 (.61, .93)	.63 (.56, .7)	.94 (.88, 1)	.93 (.9, .97)	.45 (.15, .76)	.97 (.94, 1)
	Comparison	.94 (.86, 1)	.87 (.79, .95)	.92 (.84, 1)	.29 (.2, .38)	.5 (.3, .71)	.55 (.17, .92)	.59 (.37, .82)	.65 (.32, .98)	.15 (0, .33)	.96 (.92, 1)
7.9 Improved sanitation	MV	.66 (.55, .77)	1 (.99, 1)	.22 (.09, .34)	.25 (.16, .33)	.72 (.58, .86)	.79 (.73, .85)	.57 (.5, .65)	.74 (.68, .8)	.86 (.8, .91)	.52 (.42, .61)
	Comparison	.72 (.52, .93)	.86 (.79, .94)	.03 (0, .06)	.12 (.04, .21)	.09 (.01, .17)	.61 (.51, .71)	.54 (.45, .63)	.32 (.22, .41)	.48 (.23, .73)	.47 (.41, .53)

Table 21 (does not correspond to any figures directly): classical estimates of MV and comparison outcomes in 2010 with 95% intervals of uncertainty for each outcome and country, presented on the raw scale of the data (rather than in standard deviations). Most outcomes are proportions and so range from 0 to 1. Exceptions are: the asset index, the first principal component of a list of assets;[9, 10] gross attendance, the ratio of total attendants in primary school to those of primary school age; gender parity, the ratio of girl to boy gross attendance ratios; and infant and under-five mortality rates, deaths per 1000 live births.

Tables 22 and 23 include estimation of change from 2010 to 2015 and 95% intervals of uncertainty, averaged across countries. As described, outcomes were standardized, reoriented, and subsequently averaged into outcome indices. Thus, results are on the scale of outcome standard deviations.

Category	Outcome	Estimate (95% uncertainty interval)
Poverty	1.1 Population below 1.25 USD/day	0.3 (0.2, 0.4)
	1.2 Poverty gap ratio	0.4 (0.2, 0.5)
	1.1a Asset index	2 (1.8, 2.1)
	8.15a Mobile phone ownership	0.8 (0.7, 1)
	Poverty index	0.9 (0.8, 1)
Agriculture	a.2 Mineral fertilizer use	0.1 (0, 0.2)
	a.4 Improved seed use	0.3 (0.2, 0.4)
	Agriculture index	0.2 (0.1, 0.3)
Nutrition	1.8 Under-5 underweight	0.3 (0.1, 0.6)
	1.8s Under-5 stunting	0.8 (0.6, 1)
	1.8w Under-5 wasting	0.3 (0, 0.6)
	Nutrition index	0.5 (0.3, 0.7)
Education	2.1n Net attendance	0.2 (0.1, 0.4)
	2.1g Gross attendance	0.5 (0.3, 0.6)
	2.2 Primary school completion	0.3 (0, 0.6)
	3.1 Gender parity	0 (-0.1, 0.2)
	b.1 Preschool attendance	0.5 (0.3, 0.7)
	b.3 Primary school intake	0.3 (0.1, 0.5)
	Education index	0.3 (0.2, 0.4)
Child health	4.1 Under-5 mortality	0.5 (0.2, 0.7)
	4.2 Infant mortality	0.5 (0.2, 0.8)
	4.3 Measles immunization	0.4 (0.2, 0.7)
	c.1 Exclusive breastfeeding	0.7 (0.5, 1)
	Child health index	0.5 (0.3, 0.7)
Maternal health	5.2 Skilled birth attendance	1 (0.8, 1.2)
	5.3(A) Any contraceptive use	0.8 (0.6, 1)
	5.3(M) Modern contraceptive use	0.7 (0.5, 0.9)
	5.5(1) Skilled antenatal care visit	0.4 (0.2, 0.6)
	5.5(4) Four antenatal care visits	1 (0.8, 1.2)
	Maternal health index	0.8 (0.7, 0.9)
HIV and malaria	6.1p Pregnant HIV testing	0.6 (0.5, 0.8)
	6.3 AIDS knowledge	0.4 (0.3, 0.5)
	6.7 Under-5 bednet use	0.5 (0.3, 0.7)
	6.7p Pregnant bednet use	0.2 (0, 0.5)
	6.7h Bednet ownership	0.2 (0.2, 0.3)
	6.7n Bednet correct use	0.4 (0.2, 0.6)
	HIV and malaria index	0.4 (0.3, 0.5)
Water and sanitation	7.8 Improved water	0 (-0.2, 0.2)
	7.9 Improved sanitation	0.9 (0.8, 1)
	Water and sanitation index	0.4 (0.3, 0.6)
	Overall index	0.5 (0.4, 0.5)

Table 22 (corresponds to Figure 25a): estimation of change from 2010 to 2015 for individual outcomes and indices, classical results.

Outcome index	Estimate (95% uncertainty interval)
Poverty index	0.9 (0.8, 1)
Agriculture index	0.2 (0.1, 0.3)
Nutrition index	0.5 (0.3, 0.7)
Education index	0.3 (0.2, 0.4)
Child health index	0.5 (0.3, 0.7)
Maternal health index	0.8 (0.7, 0.9)
HIV and malaria index	0.4 (0.3, 0.5)
Water and sanitation index	0.4 (0.3, 0.6)
Overall index	0.5 (0.4, 0.5)

Table 23 (corresponds to Figure 25b): estimation of change from 2010 to 2015 for outcome indices, classical results.

Outcome	Year	Estimate (95% uncertainty interval) Senegal	Mali	Ghana	Nigeria	Ethiopia	Uganda	Kenya	Rwanda	Tanzania	Malawi
1.1 Population below 1.25 USD/day	2010	.35 (.23, .47)	.93 (.89, .97)	.18 (.08, .28)	.8 (.71, .88)	.84 (.79, .88)	.41 (.32, .5)	.86 (.8, .91)	.89 (.84, .93)	.88 (.84, .93)	.82 (.76, .88)
	2015	.15 (.06, .24)	.94 (.91, .98)	.02 (0, .04)	.81 (.73, .89)	.72 (.62, .81)	.47 (.41, .53)	.77 (.7, .83)	.71 (.62, .8)	.75 (.59, .91)	.71 (.66, .77)
1.2 Poverty gap ratio	2010	.09 (.04, .14)	.53 (.46, .59)	.05 (.02, .08)	.38 (.33, .44)	.28 (.26, .3)	.12 (.1, .14)	.47 (.44, .51)	.48 (.43, .54)	.43 (.37, .49)	.37 (.34, .41)
	2015	.03 (.01, .04)	.51 (.47, .55)	.01 (0, .02)	.4 (.34, .46)	.22 (.17, .28)	.13 (.1, .16)	.34 (.31, .38)	.31 (.25, .37)	.33 (.21, .44)	.28 (.24, .31)
1.1a Asset index	2010	-1.2 (-1.89, -.5)	-2.2 (-2.62, -1.79)	-2.23 (-2.55, -1.92)	.1 (-.34, .54)	-5.17 (-5.3, -5.04)	-.77 (-1.08, -.46)	-3.35 (-3.64, -3.06)	.15 (-.55, .84)	-.92 (-1.26, -.58)	.05 (-.22, .32)
	2015	-.24 (-.79, .32)	.98 (.41, 1.55)	1.3 (.87, 1.73)	.18 (-.19, .55)	2.91 (2.79, 3.02)	.42 (-.14, .98)	2.21 (2.09, 2.33)	.89 (.36, 1.41)	1.04 (.76, 1.31)	.27 (-.21, .75)
8.15a Mobile phone ownership	2010	.94 (.9, .99)	.71 (.61, .82)	.53 (.39, .67)	.64 (.54, .74)	.02 (0, .03)	.46 (.39, .54)	.61 (.55, .66)	.64 (.54, .74)	.53 (.51, .56)	.43 (.36, .49)
	2015	.92 (.87, .97)	.89 (.82, .96)	.85 (.78, .92)	.73 (.66, .79)	.42 (.36, .48)	.63 (.53, .73)	.8 (.75, .84)	.74 (.67, .8)	.75 (.7, .8)	.43 (.38, .49)
a.2 Mineral fertilizer use	2010	.91 (.79, 1)	.46 (.33, .59)	.39 (.31, .48)	.99 (.98, 1)	.81 (.73, .88)	0 (0, 0)	.76 (.69, .82)	.16 (0, .37)	.86 (.79, .93)	.97 (.95, .99)
	2015	.93 (.84, 1)	.48 (.35, .6)	.47 (.36, .58)	.97 (.93, 1)	.97 (.95, .99)	0 (0, 0)	.84 (.8, .89)	.04 (.01, .07)	.92 (.86, .98)	.98 (.96, 1)
a.4 Improved seed use	2010	.95 (.87, 1)	.01 (0, .03)	.07 (.01, .13)	.1 (.05, .15)	.03 (.01, .06)	0 (0, 0)	.46 (.37, .55)	.07 (.02, .13)	.52 (.34, .71)	.88 (.86, .91)
	2015	.95 (.87, 1)	.39 (.29, .49)	.31 (.2, .43)	.27 (.16, .37)	.08 (.03, .13)	.01 (0, .03)	.49 (.39, .6)	.07 (.04, .1)	.61 (.5, .71)	.88 (.83, .93)
1.8 Under-5 underweight	2010	.19 (.1, .29)	.23 (.17, .3)	.09 (.05, .13)	.28 (.19, .37)	.36 (.25, .47)	.12 (.08, .16)	.21 (.09, .32)	.07 (.03, .11)	.1 (.05, .14)	.12 (.07, .17)
	2015	.11 (.06, .17)	.31 (.21, .41)	.16 (0, .33)	.02 (0, .04)	.25 (.16, .34)	.08 (.04, .13)	.02 (0, .06)	.05 (.01, .09)	.09 (.06, .13)	.19 (.1, .29)
1.8s Under-5 stunting	2010	.25 (.18, .32)	.49 (.43, .55)	.23 (.16, .3)	.5 (.38, .61)	.63 (.54, .73)	.51 (.46, .56)	.33 (.17, .5)	.26 (.16, .37)	.37 (.25, .49)	.48 (.43, .54)
	2015	.18 (.12, .24)	.43 (.31, .54)	.14 (.05, .23)	.03 (.01, .06)	.46 (.38, .55)	.42 (.36, .48)	.11 (.02, .2)	.15 (.06, .23)	.2 (.12, .28)	.46 (.39, .53)
1.8w Under-5 wasting	2010	.11 (.03, .19)	.02 (0, .04)	.02 (0, .05)	.11 (.05, .17)	.1 (.04, .15)	.03 (.01, .05)	.15 (.04, .26)	.03 (0, .07)	.06 (.01, .1)	.06 (.02, .1)
	2015	.12 (.03, .2)	.13 (.05, .22)	.03 (0, .06)	.03 (0, .07)	.06 (.01, .11)	0 (0, 0)	0 (0, 0)	.02 (0, .04)	.03 (0, .06)	.03 (0, .05)
2.1n Net attendance	2010	.5 (.36, .65)	.43 (.3, .57)	.65 (.5, .81)	.49 (.37, .62)	.49 (.41, .57)	.73 (.66, .81)	.82 (.75, .9)	.95 (.91, .99)	.83 (.81, .86)	.94 (.9, .97)
	2015	.52 (.39, .65)	.56 (.47, .65)	.81 (.72, .9)	.53 (.4, .66)	.93 (.89, .96)	.73 (.68, .78)	.75 (.71, .79)	.9 (.85, .96)	.83 (.81, .84)	.9 (.87, .93)
2.1g Gross attendance	2010	.63 (.47, .8)	.58 (.4, .75)	.94 (.72, 1.16)	.76 (.6, .92)	.85 (.75, .95)	1.12 (1, 1.25)	1.2 (1.08, 1.31)	1.35 (1.21, 1.49)	1.15 (1.12, 1.17)	1.52 (1.46, 1.59)
	2015	.96 (.47, 1.44)	.79 (.65, .93)	1.36 (1.18, 1.53)	.92 (.71, 1.13)	1.46 (1.32, 1.59)	1.19 (1.1, 1.27)	1.4 (1.25, 1.55)	1.49 (1.31, 1.67)	1.19 (1.11, 1.27)	1.66 (1.56, 1.76)
2.2 Primary school completion	2010	.4 (.18, .61)	.67 (.52, .82)	.99 (.98, 1)	.85 (.64, 1)	.26 (.13, .39)	.8 (.66, .94)	.5 (.33, .68)	.74 (.56, .92)	.82 (.7, .94)	0 (0, .01)
	2015	.73 (.57, .89)	.42 (.29, .56)	.93 (.84, 1)	.97 (.89, 1)	.67 (.45, .89)	.67 (.5, .83)	.74 (.51, .96)	.71 (.49, .92)	.89 (.77, 1)	.15 (.04, .27)
3.1 Gender parity	2010	1.15 (.73, 1.57)	1.06 (.79, 1.32)	.95 (.8, 1.1)	.76 (.6, .92)	1.3 (.97, 1.63)	1.19 (.89, 1.48)	1 (.81, 1.18)	1.11 (.86, 1.36)	1.01 (.99, 1.03)	1.03 (.85, 1.21)
	2015	1.17 (.81, 1.53)	.95 (.78, 1.11)	1.14 (.99, 1.3)	1.55 (.84, 2.25)	.8 (.67, .94)	1.16 (.97, 1.36)	1.03 (.8, 1.27)	1.07 (.9, 1.25)	.99 (.85, 1.12)	.96 (.81, 1.11)
b.1 Preschool attendance	2010	.44 (.32, .57)	.06 (.01, .1)	.61 (.5, .73)	.18 (.1, .26)	.04 (.01, .08)	.42 (.38, .47)	.44 (.34, .55)	.28 (.18, .38)	.17 (.07, .27)	.23 (.18, .27)
	2015	.35 (.25, .44)	.34 (.22, .45)	.78 (.61, .96)	.43 (.3, .56)	.11 (.05, .16)	.66 (.56, .75)	.46 (.23, .69)	.57 (.4, .74)	.17 (.03, .31)	.56 (.44, .69)
b.3 Primary school intake	2010	.09 (.02, .15)	.14 (.06, .22)	.19 (.05, .33)	.12 (.03, .21)	.02 (0, .06)	.25 (.15, .35)	.24 (.04, .44)	.61 (.48, .75)	.36 (.29, .43)	.21 (.14, .27)
	2015	.18 (.07, .3)	.21 (.1, .32)	.14 (0, .33)	.19 (.07, .32)	.69 (.54, .84)	.22 (.1, .34)	.17 (0, .36)	.33 (.18, .48)	.39 (.28, .51)	.6 (.46, .73)
4.1 Under-5 mortality	2010	63 (33, 92)	270 (227, 313)	60 (34, 86)	102 (59, 146)	57 (30, 85)	71 (32, 110)	115 (49, 181)	23 (0, 47)	69 (42, 95)	123 (46, 200)
	2015	33 (8, 58)	190 (130, 250)	6 (-3, 14)	48 (21, 74)	39 (15, 62)	61 (33, 89)	42 (17, 66)	37 (3, 71)	64 (42, 85)	99 (63, 136)
4.2 Infant mortality	2010	53 (26, 81)	150 (110, 191)	56 (28, 83)	74 (35, 113)	59 (30, 88)	57 (29, 84)	65 (20, 111)	17 (-10, 44)	54 (43, 65)	63 (28, 98)
	2015	34 (8, 60)	103 (73, 132)	3 (-3, 8)	24 (9, 39)	41 (17, 66)	46 (17, 75)	37 (16, 59)	7 (-6, 19)	49 (24, 75)	73 (31, 115)
4.3 Measles immunization	2010	.42 (.26, .59)	.43 (.3, .57)	.92 (.83, 1)	.39 (.23, .55)	.72 (.53, .9)	.65 (.53, .76)	.79 (.57, 1)	.71 (.42, 1)	.76 (.64, .89)	.9 (.82, .99)
	2015	.83 (.7, .96)	.88 (.7, 1)	.67 (.42, .92)	.75 (.59, .91)	.9 (.8, 1)	.6 (.37, .82)	.77 (.54, .99)	.84 (.63, 1)	.67 (.58, .77)	.96 (.87, 1)
c.1 Exclusive breastfeeding	2010	.65 (.46, .85)	.89 (.77, 1)	.62 (.36, .89)	.64 (.46, .83)	0 (0, 0)	.87 (.77, .97)	0 (0, 0)	.84 (.56, 1)	.31 (.16, .46)	.85 (.72, .98)
	2015	.85 (.71, 1)	.89 (.77, 1)	.66 (.42, .9)	.75 (.56, .94)	.76 (.52, 1)	.92 (.78, 1)	1 (1, 1)	.65 (.51, .79)	.66 (.5, .83)	.92 (.76, 1)
5.2 Skilled birth attendance	2010	.88 (.78, .98)	.54 (.39, .68)	.8 (.66, .95)	.17 (.07, .27)	.07 (.01, .14)	.8 (.72, .87)	.7 (.52, .88)	1 (1, 1)	.87 (.82, .92)	.88 (.82, .95)
	2015	.97 (.92, 1)	.78 (.63, .92)	.9 (.76, 1)	.81 (.71, .91)	.99 (.97, 1)	.88 (.79, .97)	.85 (.66, 1)	.97 (.93, 1)	.96 (.96, .97)	.97 (.94, 1)
5.3(A) Any contraceptive use	2010	.15 (.06, .23)	.27 (.17, .36)	.44 (.32, .55)	.26 (.18, .33)	.41 (.35, .47)	.39 (.3, .48)	.47 (.4, .54)	.51 (.35, .67)	.45 (.37, .53)	.61 (.53, .68)
	2015	.39 (.3, .47)	.51 (.39, .64)	.49 (.39, .59)	.74 (.64, .84)	.6 (.53, .68)	.72 (.6, .84)	.74 (.66, .82)	.61 (.37, .84)	.45 (.26, .64)	.75 (.69, .82)
5.3(M) Modern contraceptive use	2010	.15 (.06, .23)	.26 (.16, .35)	.35 (.27, .43)	.23 (.16, .31)	.42 (.35, .48)	.34 (.27, .42)	.43 (.38, .48)	.45 (.29, .61)	.43 (.35, .52)	.59 (.51, .68)
	2015	.38 (.29, .47)	.39 (.28, .5)	.43 (.3, .55)	.7 (.6, .81)	.53 (.46, .6)	.59 (.48, .7)	.73 (.66, .81)	.6 (.43, .78)	.43 (.25, .61)	.73 (.66, .81)
5.5(1) Skilled antenatal care visit	2010	.97 (.94, 1)	.68 (.51, .84)	.93 (.83, 1)	.75 (.63, .87)	.85 (.75, .96)	.99 (.98, 1)	.99 (.97, 1)	1 (1, 1)	.96 (.91, 1)	.98 (.95, 1)
	2015	1 (1, 1)	.94 (.87, 1)	.92 (.82, 1)	1 (.99, 1)	.99 (.97, 1)	.93 (.86, 1)	.97 (.92, 1)	.96 (.91, 1)	.95 (.92, .98)	.99 (.96, 1)
5.5(4) Four antenatal care visits	2010	.5 (.35, .65)	.25 (.18, .31)	.86 (.76, .97)	.58 (.49, .67)	.55 (.44, .65)	.83 (.75, .9)	.58 (.4, .77)	.45 (.29, .62)	.6 (.51, .69)	.4 (.31, .49)
	2015	.84 (.74, .94)	.76 (.66, .86)	.86 (.73, .99)	.97 (.94, 1)	.88 (.77, .98)	.87 (.79, .95)	.75 (.62, .88)	.64 (.47, .81)	.93 (.88, .97)	.49 (.31, .68)
6.1p Pregnant HIV testing	2010	.56 (.42, .69)	.44 (.3, .57)	.61 (.45, .77)	.57 (.43, .71)	.91 (.85, .98)	.9 (.84, .96)	.97 (.93, 1)	.98 (.94, 1)	.91 (.86, .96)	.96 (.93, .99)
	2015	.87 (.82, .93)	.83 (.73, .93)	.8 (.63, .97)	1 (.99, 1)	.99 (.98, 1)	.99 (.96, 1)	1 (1, 1)	1 (1, 1)	.98 (.96, 1)	.99 (.96, 1)
6.3 AIDS knowledge	2010	.09 (.04, .14)	.03 (.02, .04)	.27 (.12, .42)	.28 (.23, .33)	.11 (.05, .17)	.39 (.35, .43)	.46 (.41, .52)	.43 (.31, .55)	.03 (0, .07)	.55 (.52, .57)
	2015	.09 (.06, .12)	.07 (.04, .11)	.14 (.1, .17)	.76 (.71, .8)	.88 (.74, 1)	.26 (.2, .32)	.62 (.58, .67)	.24 (.19, .29)	.21 (.19, .24)	.61 (.56, .67)
6.7 Under-5 bednet use	2010	.45 (.28, .61)	.57 (.51, .63)	.54 (.41, .66)	.68 (.57, .78)	.56 (.48, .65)	.43 (.36, .49)	.56 (.4, .73)	.56 (.46, .65)	.78 (.73, .84)	.68 (.63, .73)
	2015	.66 (.56, .76)	.76 (.66, .86)	.51 (.33, .7)	.75 (.66, .84)	.61 (.44, .77)	.62 (.52, .73)	.73 (.61, .85)	.72 (.67, .76)	.73 (.59, .88)	.88 (.84, .93)
6.7p Pregnant bednet use	2010	.46 (.18, .73)	.71 (.56, .86)	.52 (.27, .77)	.74 (.58, .9)	.77 (.59, .96)	.64 (.41, .87)	.43 (.16, .71)	.58 (.19, .98)	.82 (.7, .94)	.33 (.14, .52)
	2015	.75 (.55, .95)	.74 (.58, .91)	.56 (.31, .8)	.79 (.61, .97)	.41 (.21, .6)	.55 (.27, .83)	.57 (.19, .95)	.81 (.56, 1)	.83 (.8, .85)	.92 (.75, 1)
6.7h Bednet ownership	2010	1 (.99, 1)	.99 (.98, 1)	.77 (.68, .86)	.97 (.95, .99)	.97 (.95, .99)	.83 (.76, .91)	.78 (.71, .84)	.85 (.78, .92)	.96 (.95, .98)	.75 (.71, .79)
	2015	.93 (.9, .96)	.97 (.94, 1)	.84 (.79, .9)	.98 (.97, 1)	.93 (.88, .98)	.97 (.95, .99)	.92 (.88, .96)	.89 (.86, .91)	.99 (.98, 1)	.96 (.94, .98)
6.7n Bednet correct use	2010	.67 (.57, .77)	0 (0, 0)	.51 (.36, .66)	.79 (.72, .86)	.45 (.37, .53)	.69 (.61, .78)	.82 (.75, .9)	.55 (.5, .61)	.63 (.57, .69)	.52 (.44, .6)
	2015	.66 (.57, .74)	.59 (.51, .66)	.66 (.58, .73)	.82 (.76, .89)	.41 (.34, .48)	.68 (.61, .75)	.76 (.7, .82)	.62 (.53, .72)	.72 (.68, .77)	.47 (.38, .55)
7.8 Improved water	2010	.75 (.6, .9)	.8 (.73, .87)	.82 (.66, .98)	.66 (.52, .8)	.87 (.78, .96)	.32 (.2, .44)	.92 (.88, .97)	.98 (.96, 1)	.64 (.36, .91)	.96 (.93, .99)
	2015	.73 (.59, .87)	.63 (.55, .71)	.94 (.84, 1)	.6 (.45, .75)	.77 (.61, .93)	.63 (.56, .7)	.94 (.88, 1)	.93 (.9, .97)	.45 (.15, .76)	.97 (.94, 1)
7.9 Improved sanitation	2010	.73 (.6, .87)	.43 (.37, .5)	.11 (.04, .18)	.11 (.06, .16)	.63 (.48, .77)	.67 (.56, .79)	.19 (.13, .25)	.52 (.35, .68)	.06 (0, .14)	.02 (0, .05)
	2015	.66 (.55, .77)	1 (.99, 1)	.22 (.09, .34)	.25 (.16, .33)	.72 (.58, .86)	.79 (.73, .85)	.57 (.5, .65)	.74 (.68, .8)	.86 (.8, .91)	.52 (.42, .61)

Table 24 (corresponds to Figures 26-33): classical estimates of project year five (2010) and ten (2015) outcomes with 95% intervals of uncertainty for each outcome and country, presented on the raw scale of the data (rather than in standard deviations). Most outcomes are proportions and so range from 0 to 1. Exceptions are: the asset index, the first principal component of a list of assets;[9, 10] gross attendance, the ratio of total attendants in primary school to those of primary school age; gender parity, the ratio of girl to boy gross attendance ratios; and infant and under-five mortality rates, deaths per 1000 live births.

References

[1] Sachs J. Rapid Victories against Extreme Poverty. Scientific American. 2007 April;296(4):34.

[2] Sachs JD, McArthur JW, Schmidt-Traub G, Kruk M, Bahadur C, McCord G. Ending Africa's Poverty Trap. Brookings Papers on Economic Activity. 2004;1:117–240. Available from: http://www.unmillenniumproject.org/documents/BPEAEndingAfricasPovertyTrapFINAL.pdf.

[3] Clemens MA, Demombynes G. When does rigorous impact evaluation make a difference? The case of the Millennium Villages. Journal of Development Effectiveness. 2011 September;3(3):305–339.

[4] UN Statistics Division. Millennium Development Goals Indicators; 2017 [cited 2015]. Available from: http://mdgs.un.org/unsd/mdg/Metadata.aspx.

[5] United Nations Development Group. Indicators for monitoring the Millennium Development Goals: Definitions, rationale, concepts and sources. New York, NY: United Nations; 2003. ST/ESA/STAT/SER.F/95.

[6] UNESCO Institute for Statistics. Global Education Digest 2010: Comparing Education Statistics Across the World. Montreal, Quebec: UNESCO Institute for Statistics; 2010. Available from: http://www.uis.unesco.org/Library/Documents/GED_2010_EN.pdf.

[7] World Health Organization. Health situation and trends assessment;. Available from: http://www.searo.who.int/entity/health_situation_trends/data/mdg/measles/en/.

[8] MVP. Study protocol, integrating the delivery of health and development interventions: assessing the impact on child survival in sub-Saharan Africa.; 2009. Available from: https://ciesin.columbia.edu/confluence/download/attachments/91488269/MVP+Evaluation+Protocol.pdf.

[9] Filmer D, Pritchett LH. Estimating Wealth Effects Without Expenditure Data - Or Tears: An Application to Educational Enrollments in States of India. Demography. 2001 February;38(1):115–132.

[10] Michelson H, Muniz M, DeRosa K. Measuring Socio-economic Status in the Millennium Villages: The Role of Asset Index Choice. The Journal of Development Studies. 2013;49(7):917–935.

[11] Foster J, Greer J, Thorbecke E. A Class of Decomposable Poverty Measures. Econometrica. 1984 May;52(3):761–766.

[12] Rutstein SO, Rojas G. Guide to DHS Statistics. Demographic and Health Surveys, Calverton, Maryland: Demographic and Health Surveys; 2006. Available from: http://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_29Oct2012_DHSG1.pdf.

[13] Mitchell S, Gelman A, Ross R, Huynh UK, McClellan L, Harris M, et al. The Millennium Villages Project: A protocol for the final evaluation. Protocol in The Lancet. 2015;Available from: http://www.thelancet.com/doi/story/10.1016/html.2015.07.03.2167.

[14] The World Bank. DataBank; 2017 [cited 2015]. Available from: http://databank.worldbank.org/data/reports.aspx?source=2&country=&series=SP.DYN.IMRT.IN&period=#.

[15] World Health Organization. Nutrition Landscape Information System (NLiS): Integrated WHO Nutrition Global Databases; 2017 [cited 2015]. Available from: http://apps.who.int/nutrition/landscape/search.aspx?dm=52&countries=.

[16] Demographic and Health Surveys. Dataset Access; 2017 [cited 2015]. Available from: http://dhsprogram.com/data.

[17] UNICEF. UNICEF Multiple Indicator Cluster Surveys (MICS); 2017 [cited 2015]. Available from: http://mics.unicef.org/.

[18] GPWv3. Socioeconomic Data and Applications Center (SEDAC): Gridded Population of the World (GPW), v3;. Available from: http://sedac.ciesin.columbia.edu/data/collection/gpw-v3 [cited October 2014].

[19] GADMv2. Global Administrative Areas Databse (GADMv2); 2012. Available from: http://www.gadm.org [cited 2015].

[20] Dixon J, Gulliver A, Gibbon D. Farming Systems and Poverty: Improving Farmers' Livelihoods in a Changing World. Rome and Washington DC: FAO and the World Bank; 2001.

[21] Joint Research Centre: Land Resource Management Unit. Travel time to major cities: A global map of Accessibility;. Available from: http://bioval.jrc.ec.europa.eu/products/gam/sources.htm [cited October 2014].

[22] ISRIC: World Soil Information. Soil property maps of Africa at 1 km;. Available from: http://www.isric.org/data/soil-property-maps-africa-1-km [cited October 2014].

[23] IRI/LDEO. IRI/LDEO Climate Data Library;. Available from: http://iridl.ldeo.columbia.edu/ [cited October 2014].

[24] The CGIAR Consortium for Spatial Information (CGIAR-CSI). SRTM 90m Digital Elevation Data;. Available from: http://srtm.csi.cgiar.org/ [cited October 2014].

[25] Ghosh M, Rao JNK. Small Area Estimation: An Appraisal. Statistical Science. 1994;9(1):55–76.

[26] Ghosh M, Natarajan K. Small Area Estimation: A Bayesian Perspective. In: Ghosh S, Dekker M, editors. Multivariate Analysis, Design of Experiments and Survey Sampling. New York, NY: Wiley; 1999. p. 69–92.

[27] Nadram B. Bayesian Generalized Linear Models for Inference About Small Areas. In: Rey D, Ghosh SK, Mallick BK, editors. Generalized Linear Models. Boca Raton: CRC Press; 2000. p. 89–109.

[28] Rao JNK. Small Area Estimation. Hoboken, New Jersey: John Wiley and Sons; 2003.

[29] Jiang J, Lahiri P. Mixed Model Prediction and Small Area Estimation. TEST. 2006;15(1):1–96.

[30] Measure DHS/ICF International. Sampling and Household Listing Manual: Demographic and Health Surveys Methodology. Measure DHS; 2012. Available from: http://www.measuredhs.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf.

[31] Zheng H, Little RJA. Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples. Journal of Official Statistics. 2003;19(2):99–117.

[32] Zheng H, Little RJA. Penalized Spline Nonparametric Mixed Models for Inference about a Finite Population Mean from Two-Stage Samples. Survey Methodology. 2004;30:209–218.

[33] Zheng H, Little RJA. Inference for the Population Total from Probability-Proportional-to-Size Samples Based on Predictions from a Penalized Spline Nonparametric Model. Journal of Official Statistics. 2005;21(1):1–20.

[34] Chen Q, Elliott MR, Little RJA. Bayesian penalized spline model-based inference for finite population propotion in unequal probability sampling. Survey Methodology. 2010 June;36(1):23–34.

[35] Si Y, Pillai NS, Gelman A. Bayesian nonparametric weighted sampling inference. Bayesian Analysis. 2015;10(3).

[36] Fay R, Herriot R. Estimates of Income for Small Places: an Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association. 1979;366:269–277.

[37] Zaslavsky AM. Sampling from a Bayesian Menu. Statistical Science. 2011;26(2):235–237.

[38] Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. Journal of the Royal Statistical Society A. 2008;171(2):481–502.

[39] Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. Statistical Science. 2010;25(1):1–21.

[40] Ho DE, Imai K, King G. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis. 2007;15:199–236.

[41] Imbens GW, Rubin DB. Causal Inference in Statistics and Social Sciences: An Introduction. New York, NY: Cambridge University Press; 2015.

[42] US Census Bureau. CSPro: Census and Survey Processing System, Version 5.0.3. Washington DC, USA: US Census Bureau, Macro International, and Serpo, S.A.; 2013. Available from: http://www.census.gov/population/international/software/cspro/.

[43] Särndal CE, Swensson B, Wretman J. Model Assisted Survey Sampling. New York: Springer-Verlag; 1992.

[44] HemoCue Worldwide. HemoCue. Angelholm, Sweden: Radiometer Group; 2014. Available from: http://www.hemocue.com/en/products/hb-301-kit.

[45] Access Bio Inc. Somerset, NJ, USA; 2012. Available from: http://www.accessbio.net/english/product/01.asp [cited October 2014].

[46] Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. American Journal of Tropical Medicine and Hygiene. 2012 February;86(2):194–198.

[47] WHO product testing. Malaria Rapid Diagnostic Test Performance: Results of WHO product testing of malaria RDTs: Round 5 (2013). World Health Organization; 2014. ISBN 978 92 4 150755 4. Available from: http://apps.who.int/iris/bitstream/10665/128678/1/9789241507554_eng.pdf.

[48] World Health Organization. Information note on recommended selection criteria for procurement of malaria rapid diagnostic tests (RDTs). World Health Organization; 2014. Available from: http://www.who.int/malaria/publications/atoz/rdt_selection_criteria/en/.

[49] UNICEF. Height/Length Measuring Boards. UNICEF;. 18. Available from: http://www.unicef.org/supply/files/Height_Length_Measuring_Boards.pdf.

[50] Su YS, Gelman A, Hill J, Yajima M. Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. Journal of Statistical Software. 2011 December;45(2):1–31.

[51] R Development Core Team. The R project for statistical computing; 2014. Available from: http://www.r-project.org/.

[52] Catterall JS. Economic evaluation of public programs. New directions for program evaluation. 1985;1985(26):99–103.

[53] Ahren P. Economic evaluation methods in community planning. Swedish Council for Building Research; 1976.

[54] Pushpangadan P. Conservation and economic evaluation of biodiversity. Pushpangadan P, Ravi K, Santhosh V, editors. India: Science Publishers Inc; 1997.

[55] Rahman ML, Alam MF. An economic evaluation of some credit programmes designed for the small farmers and landless poor in Bangladesh. Dhaka. Bangladesh Agricultural University: Bureau of Socioeconomic Research and Training; 1987.

[56] Hutchinson BG. The economic evaluation of urban transportation investments. London Centre for Environmental Studies; 1969.

[57] Lohr SL. Sampling: Design and Analysis. 2nd ed. Cengage Learning; 2010.

[58] Lumley T. `survey`: analysis of complex survey samples; 2016. R package version 3.31-5.

[59] Collett D. Modelling Survival Data in Medical Research. Second edition ed. CRC Press; 2003.

[60] Gelman A, Hill JL. Data Analysis Using Regression and Multilevel/Hierarchical Models. New York, NY: Cambridge University Press; 2007.

[61] Gelman A. Multilevel (Hierarchical) Modeling: What it can and cannot do. Technometrics. 2006;48(3):432–435.

[62] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. Chapman & Hall/CRC texts in statistical science; 2014.

[63] Altonji JG, Elder TE, Taber CR. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy. 2005 February;113(1):151–184.