Catalyzing Inclusive Agricultural Transformation in Africa

A Machine Learning Approach

Sam Fraiberger (WB)
Marelize Görgens (WB)
Clara Ivanescu (WB)
Andrew Longosz (WB)
Shaffiq Somani (WB)
Tushar Malik (WB)
Theo Hawkins (WB)
Lakshmi Subramanian (WB/NYU)
Ananth Balashankar (NYU)
Eric Deregt (NYU)
David Wilson (BMGF)

WORLD BANK GROUP

Key Questions concerning agricultural transformation

AGRICULTURAL TRANSFORMATION

Use agricultural transformation inputs to define clusters of households of farmers that are associated with differences in productivity and income growth ☑

Are clusters consistent over time? ☑ YES

How can agricultural transformation within a cluster be optimized? ☑ DETERMINED

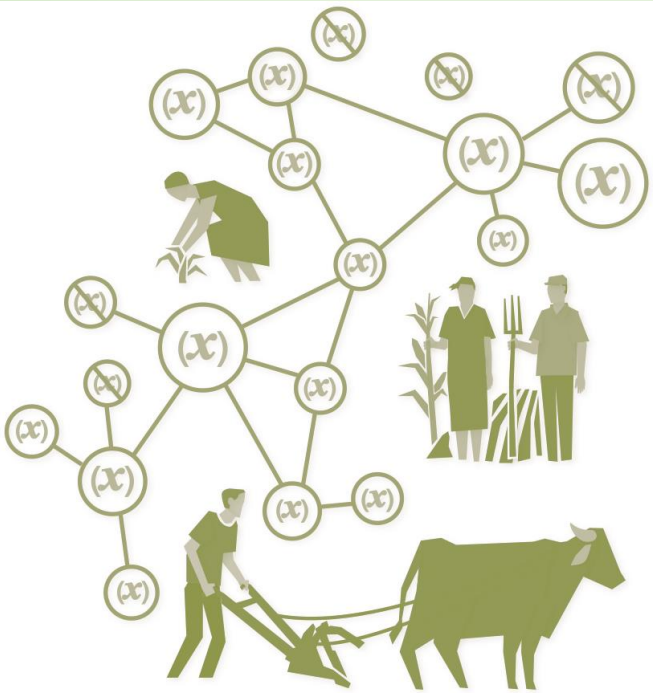
Are there pathways for progress between clusters? ☑ YES

Additional, not requested by Ag PST:

Do these differ within and between countries (Ethiopia and Tanzania)? ☑ YES

What have we learned?

- Distinct set of actionable and non-actionable variables exist that are most correlated with agricultural transformation and social service outcomes
- Input variables are correlated with one another, forming distinct groups (clusters) of farmers with similar profiles
- Can use machine learning to identify variables that are most predictive of increase in income, and predict increases in income within and across clusters



- Can identify what is most important to move between clusters and how changes to specific input variables changes the probability of moving between clusters

What have we learned?

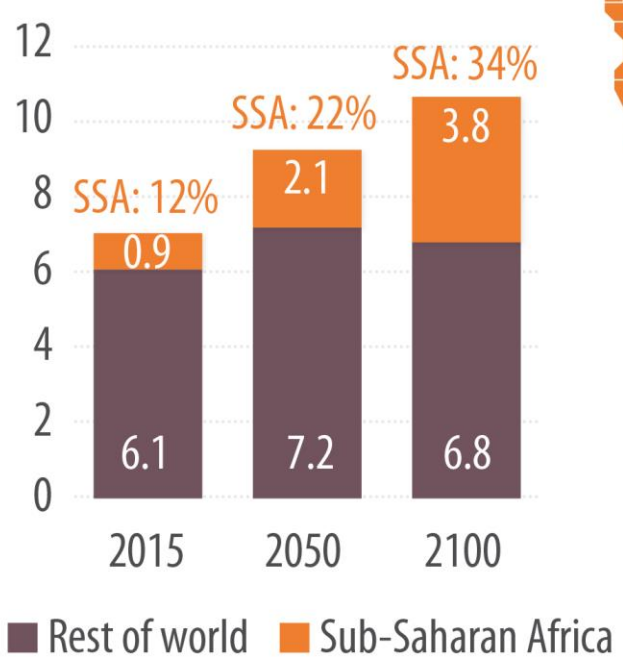
- Can identify how to maximize the income of a farmer within a cluster
- Results between countries show stark differences:
  - Different types of clusters
  - Different types of input variables
  - Different levels of correlation
  - Different predictions of the level of transformational outcomes that changes in specific input variables will bring

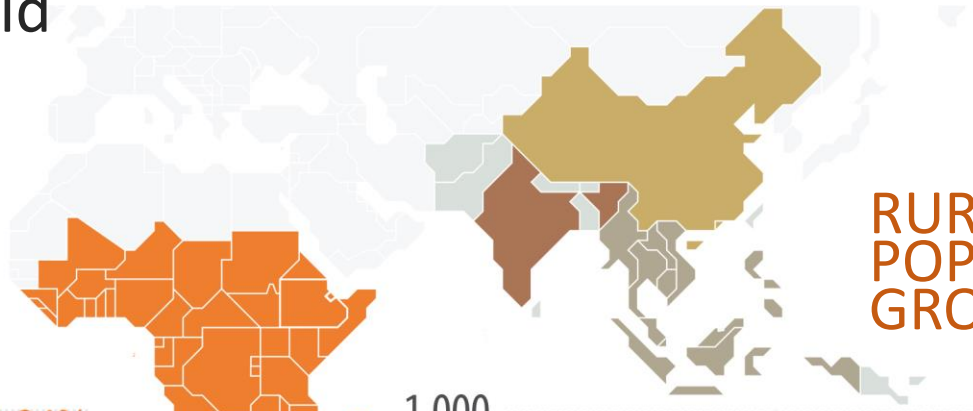
- Potential for results to be used in programmes to:
  - Develop a country-relevant taxonomy of farmers
  - Prioritize different types of support to different groups of farmers
  - Focus innovations on most important input variables, and
  - Change how results are measured and performance incentivised

Development context in Africa is rapidly changing…

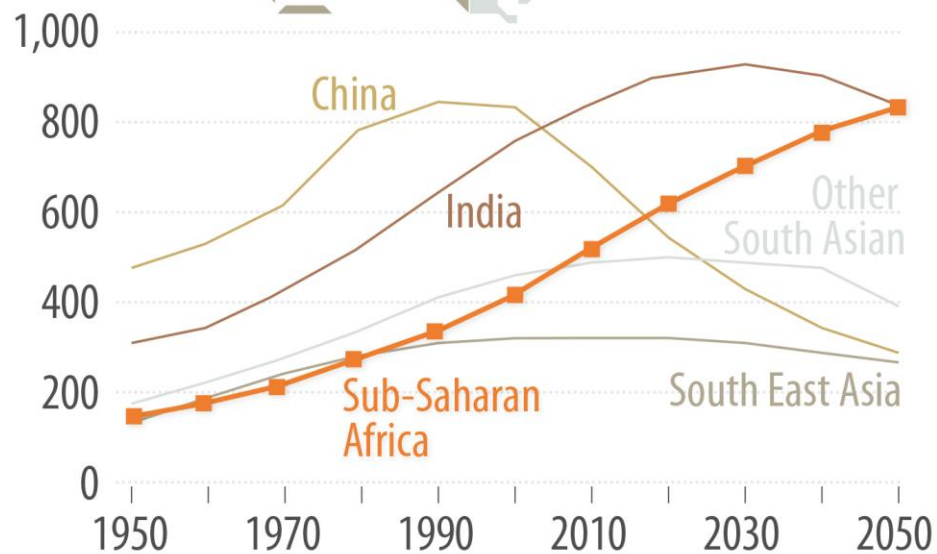
…increases in overall and rural populations unlike in other parts of the world

SSA POPULATION GROWTH





RURAL POPULATION GROWTH

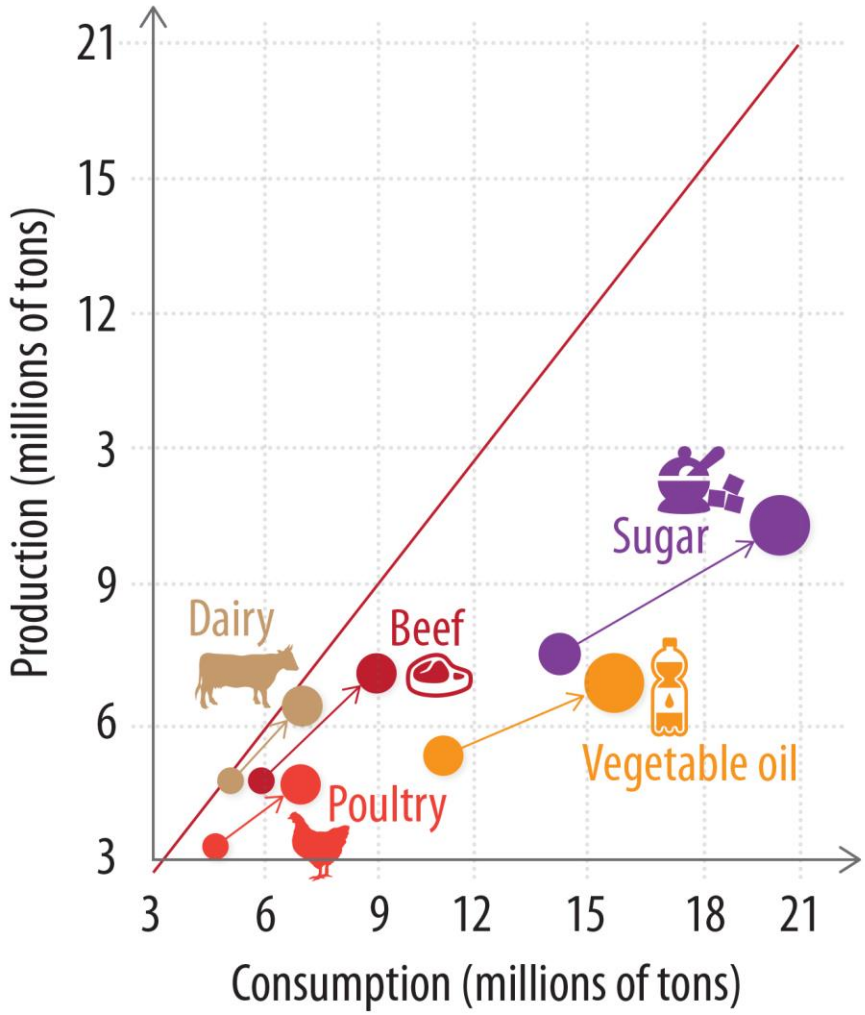


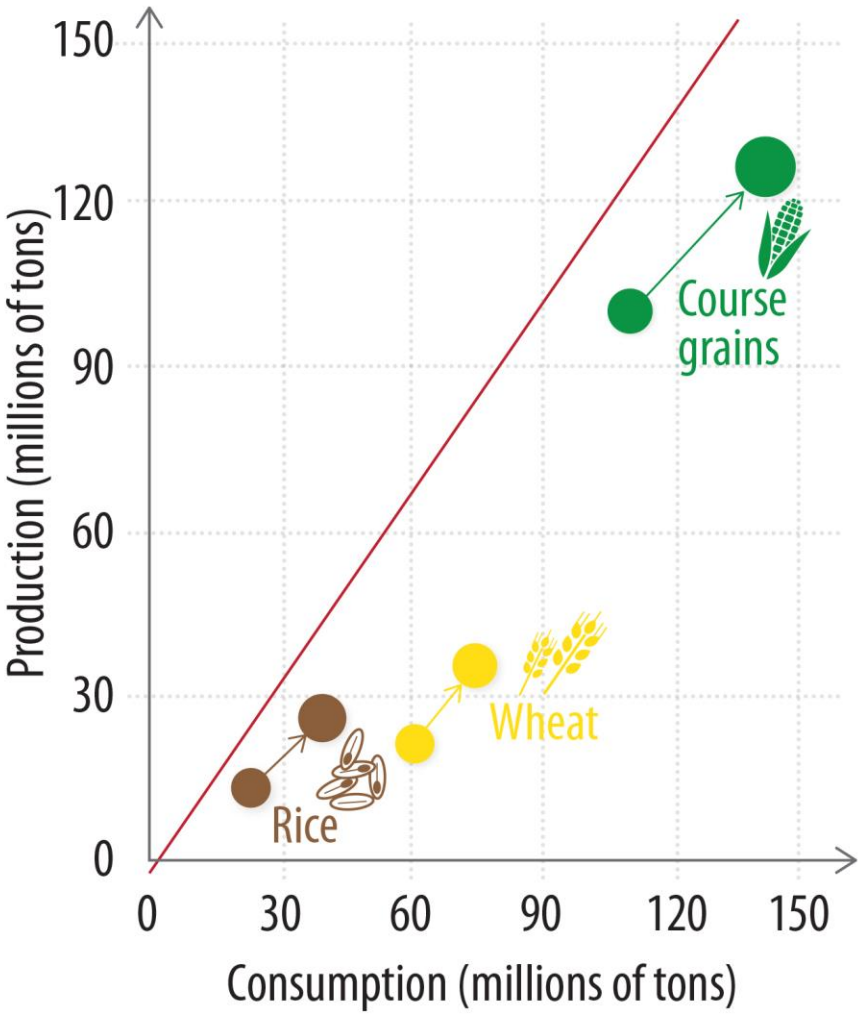
WORLD BANK GROUP

Source: Jayne et al. (2016)

Because of population growth, increased need for food in Africa

Projected trends in sub-Saharan African commodity production and consumption (2013–2023)





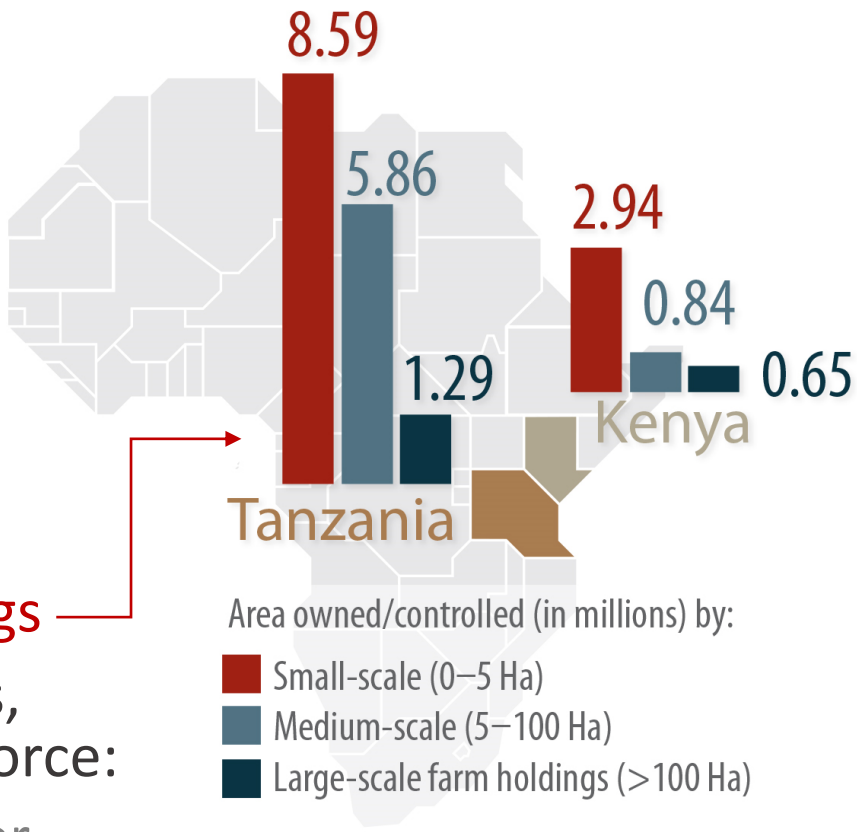
WORLD BANK GROUP

Source: FAO: acknowledgements to Holger Matthey/FAO, August 2014

Despite some transformation, agriculture in Africa struggles

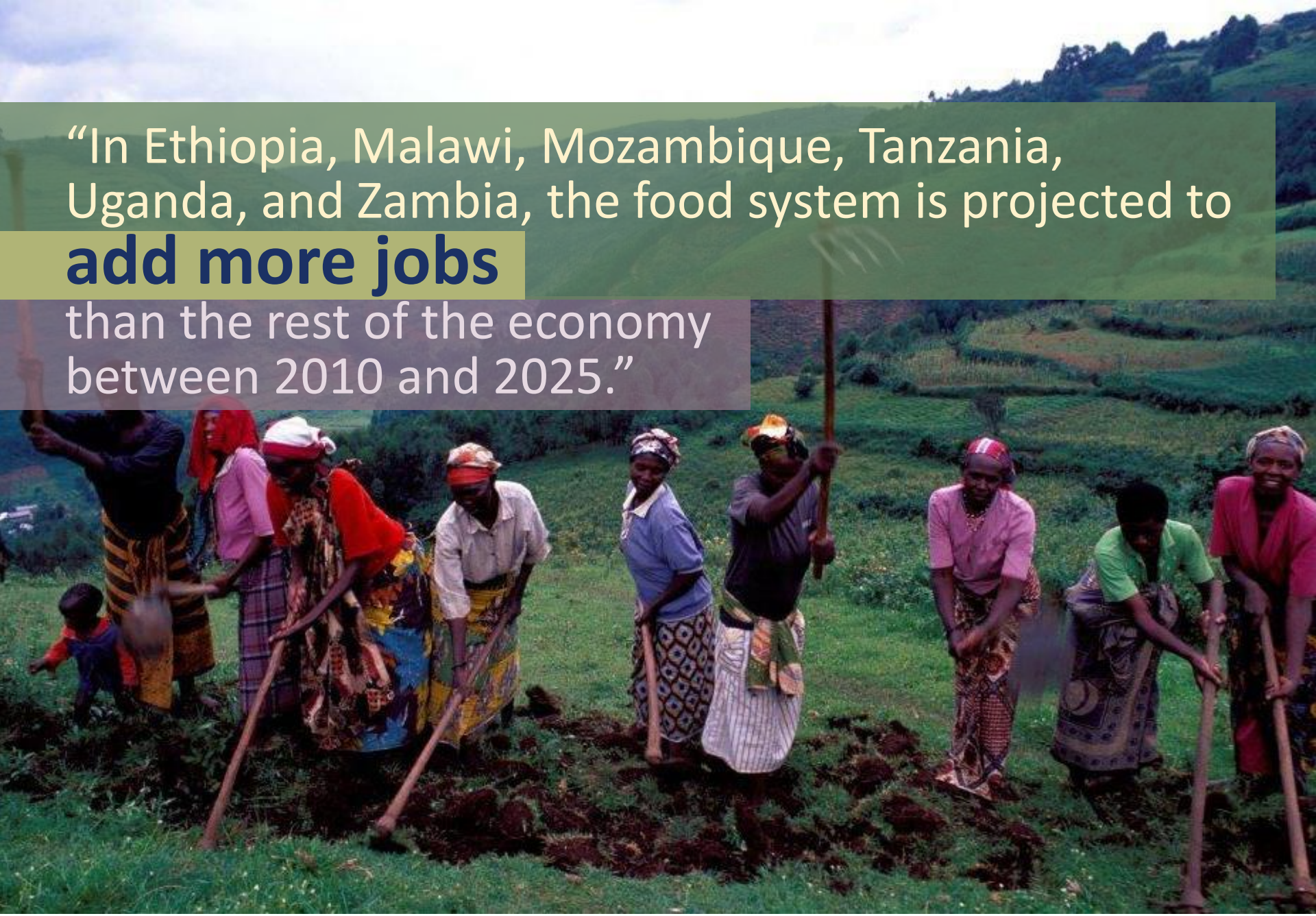
Agriculture remains the predominant sector of the economy 25% of GDP in SSA

- Yet, Africa has poorest, most unequal societies
- Most food insecure continent with high malnutrition
- Low levels of agricultural productivity and a worsening food trade balance
- Still high levels of subsistence agriculture with small landholdings
- Urgent need for 21st century jobs, yet a shrinking agriculture labor force:
  - Farming is primary employment for approx. 60% of Africa's working-age population, down from 80% a decade ago



Source: Jayne et al. (2016)

WORLD BANK GROUP



"In Ethiopia, Malawi, Mozambique, Tanzania, Uganda, and Zambia, the food system is projected to **add more jobs** than the rest of the economy between 2010 and 2025."

WORLD BANK GROUP

To achieve agriculture's potential, transformation is essential

Measured through:

- Increases in farmers' income, competitiveness and productivity
- Better food security
- Better access to social services (education and health)

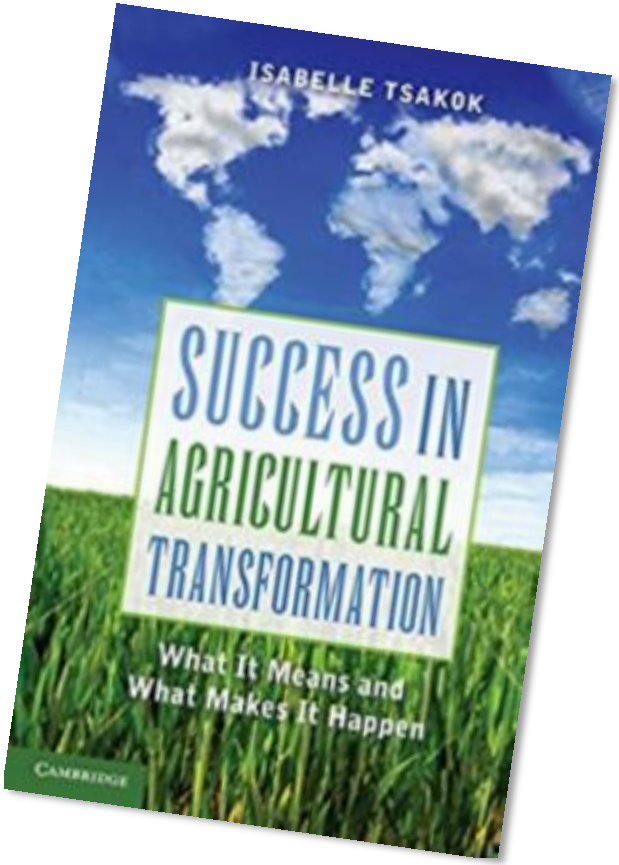
Stronger agricultural growth facilitates human capital growth and economic growth

Multiple theories about agricultural transformation pathways

2011 book by Tsakok (WB) suggests 5 conditions at the system level as necessary for successful agricultural transformation

- Stable framework of macroeconomic and political stability
- Effective technology transfer system
- Access to lucrative markets
- Ownership system
- Employment-creating non-agricultural sectors

IS THIS TRUE?

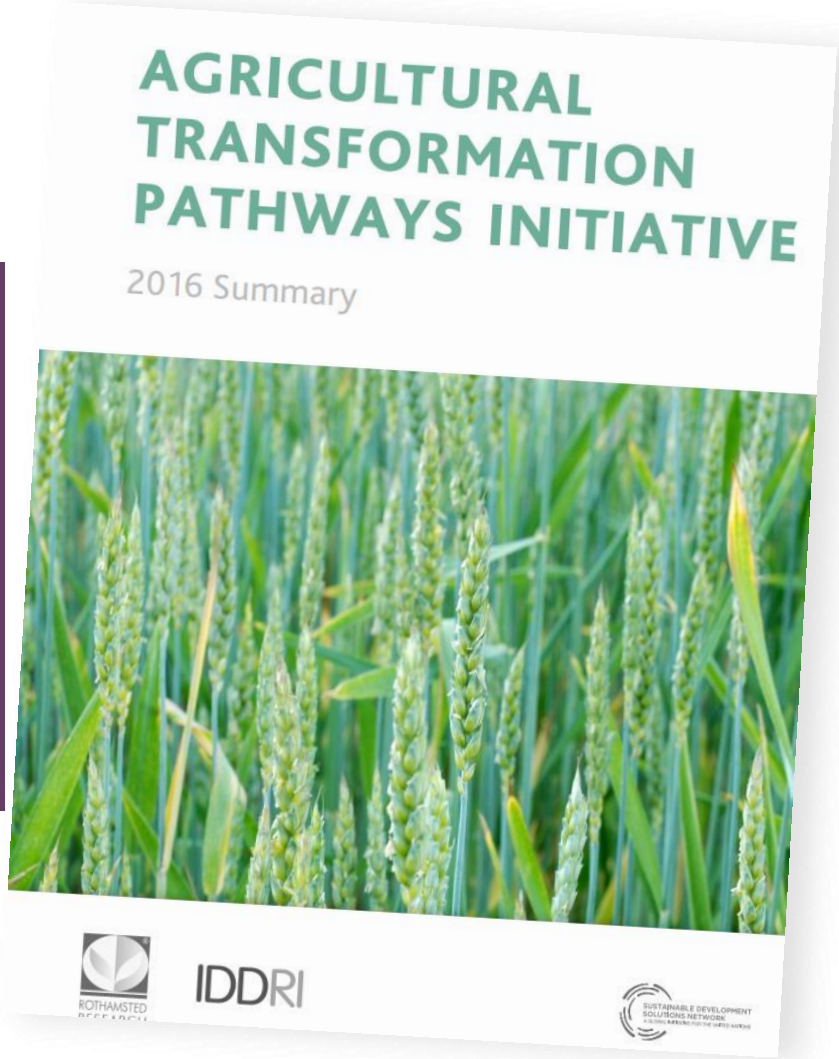


WORLD BANK GROUP

Multiple theories about agricultural transformation pathways

2016 IDDRI report suggests that pathways are unique and country-specific

IS THIS TRUE?

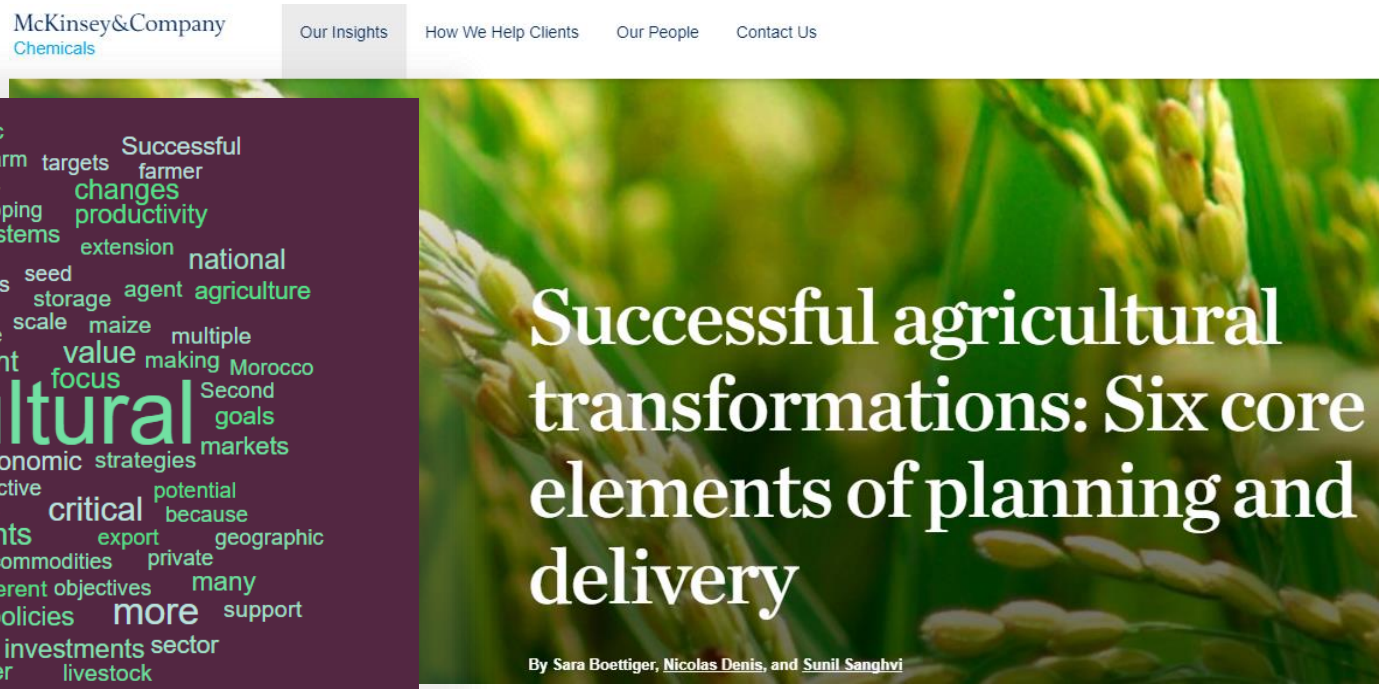


WORLD BANK GROUP

Multiple theories about agricultural transformation pathways

McKinsey (2017) suggests that there are six steps for every country to follow in the recipe towards agricultural transformation

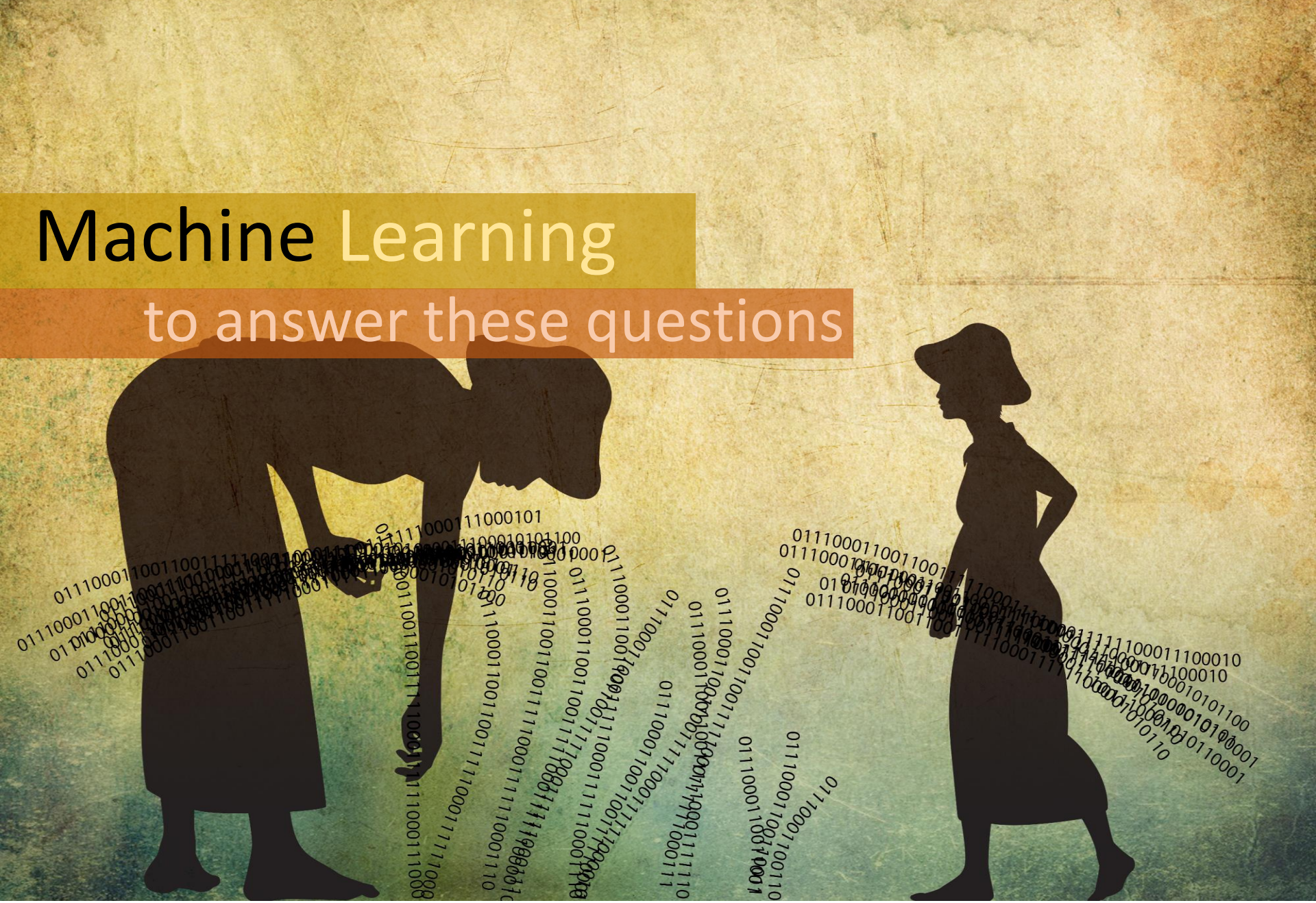
IS THIS TRUE?



WORLD BANK GROUP

Machine Learning

to answer these questions



WORLD BANK GROUP

Used the LSMS-ISA dataset

- Longitudinal survey of farmers; links farm and non-farm activities
- BMGF funding for its implementation
- 8 Countries:
  - Burkina Faso (1 wave)
  - Malawi (2 waves)
  - Niger (2 waves)
  - Tanzania (4 waves)
  - Ethiopia (3 waves)
  - Mali (1 wave)
  - Nigeria (3 waves)
  - Uganda (4 waves)

- Initial focus on Ethiopia:
  - ~3,500 households surveyed over time (2011–12, 2013–14, 2015–16)
  - ~1,500 features per households

WHY ETHIOPIA FIRST?

- Priority for Ag PST
- 3 waves of data
- Real commitment to agricultural transformation (Ethiopian ATA established)

Ethiopian ATA
Agricultural Transformation Agency
የኢትዮጵያ ግብርና ትራንስፎርሜሽን ኤጀንሲ

- Same approach expanded to Uganda and Tanzania to assess differences between countries

WORLD BANK GROUP

Limitations of LSMS-ISA dataset relevant to this dataset

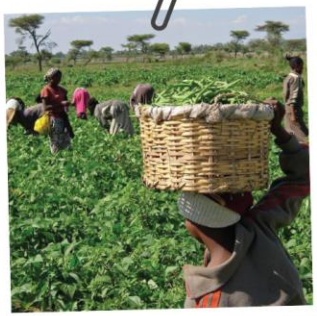
1. Zeros and missing values make it difficult to draw inferences at the subpopulation level:
   - Remove features with less than 5% of non-zero observations
   - Remove features with more than 30% missing values

2. Inconsistencies in survey questions across years /countries: Remove features that are not observed across years
3. "Small Data": Constraints the number of feasible subpopulations
4. GPS coordinates not available (Uganda)
   - Where GPS coordinates of households in panel data set are available, households are clustered together with one scrambled set of GPS coordinates per cluster (Tanzania)

What can we measure from these data?

OUTCOMES:

- Evidence of agricultural transformation and how they change over time
  - Crop sales, crop sales growth, productivity, household expenditure, food expenditure diversification, and food security
  - Education and health service access



INPUTS…

- through which to achieve agricultural transformation and how they change over time: Household, farmer and farming practices characteristics
  - Some inputs can be modified through short term policy actions (actionable) and others not (non-actionable):

ACTIONABLE

- Accessibility (distance to road/market/population center)
- Agronomic practices (crop diversification, fertilizer, seeds type, irrigation, damage prevention, land certificate, extension program)
- Equipment (axe, oxen, plough, sickle)
- Rented factors (credit, hired labor)
- Shocks (health issues, unexpected price changes)
- Financial inclusion (access to credit, bank accounts and savings)

NON-ACTIONABLE

- Demographics
- (age, marital status, region of origin)
- Physical conditions (elevation, temperature, precipitation, rooting conditions, variations in greenness)

Machine Learning Results: Ethiopia

WORLD BANK GROUP

First, unsupervised learning results and change in approach

- **INITIAL APPROACH**: Optimal segmentation of farmers into clusters was done by segmenting the population into sub-groups. Sub-groups were chosen in such a way that it yielded the highest gain with respect to population baseline in accuracy of predicting farmers' crop sales.
  - Unsupervised learning does increase the model's predictive power by 10%–15%, BUT variables most predictive of the cluster in which a farmer are non-actionable: *Precipitation, level of greenness and lives In Amhara*
  - As we wanted to develop policy recommendations, chose to focus on actionable variables for which policy interventions can be designed
- **DECIDED**: Need a **semi-supervised ML approach**
  a. Look at correlation between outcomes: are they cross-correlated to determine if outcomes should be measured separately or together
  b. Look at correlation between outcomes and input variables
  c. Identify highly-correlated input variables
  d. Cluster farmers using k-mean clustering

     In k-means clustering: Finds groups of farmers such that the values of the farmers across the 7 selected input variables are similar to others in the group and different to farmers in other clusters, i.e., minimize Euclidian distance to the centre and maximize distance between groups.

     Additional step: Weight each input by its average correlation across outcomes variables
  e. Look at most important variable/s within each cluster
  f. Look at pathways and thresholds to move between clusters

Are agric. transformation outcomes in Ethiopia correlated with each other?

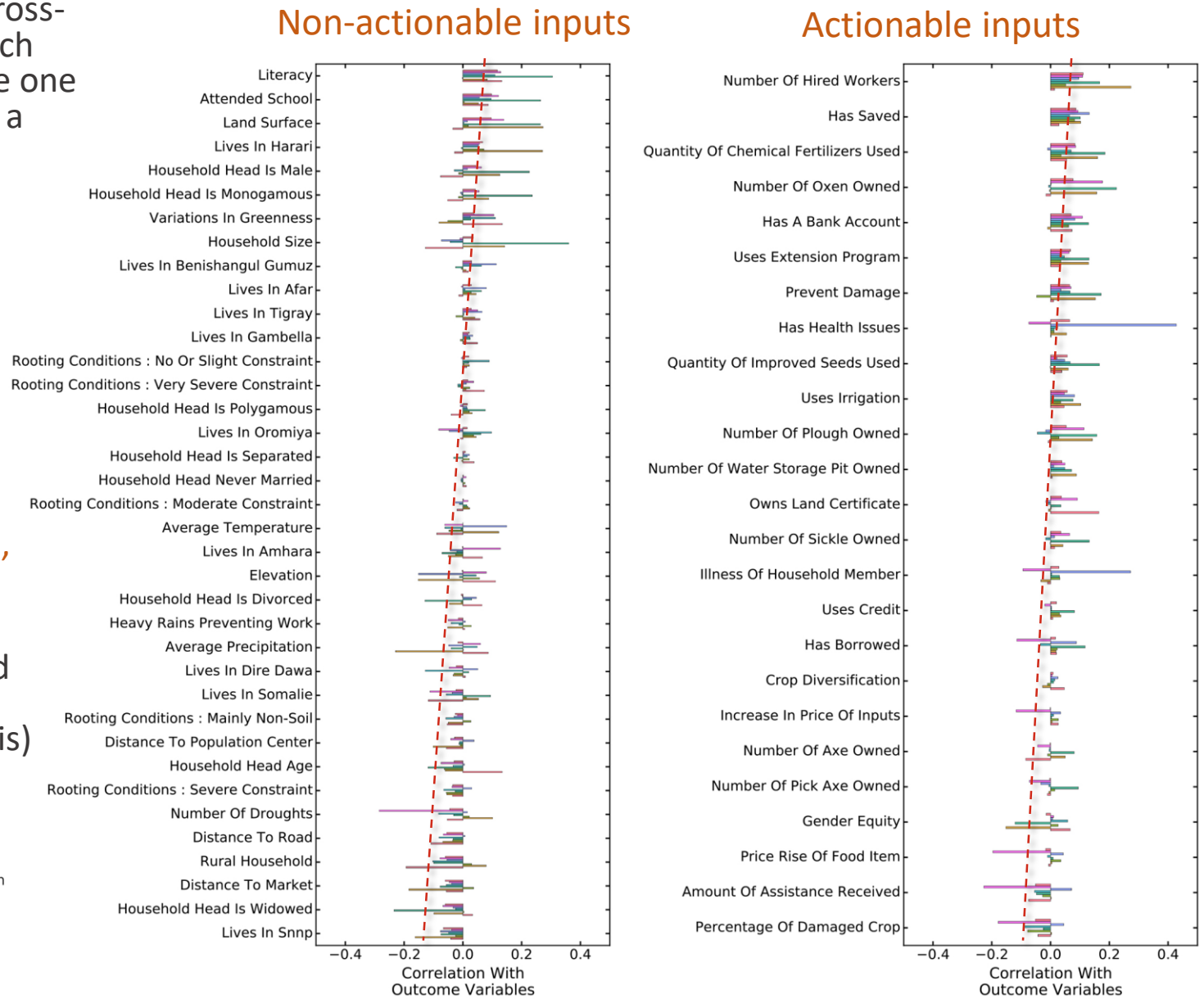
	Children Education	Crop Sales	Crop Sales Growth	Expenditure	Food Expenditure Diversification	Has Medical Assistance	No Food Deficiency
Children Education		0.011	-0.044	0.141	0.115	0.054	0.108
Crop Sales			0.45	0.273	0.047	0.062	0.174
Crop Sales Growth				0.008	-0.032	-0.023	0.043
Expenditure					0.074	0.068	0.228
Food Expenditure Diversification						0.086	0.09
Has Medical Assistance							0.005
No Food Deficiency							

- Varying levels of correlation between outcomes: mostly low
- So, need to evaluate each outcome separately in terms of its correlation with inputs

WORLD BANK GROUP

First, determine cross-correlation between inputs and selected outcomes

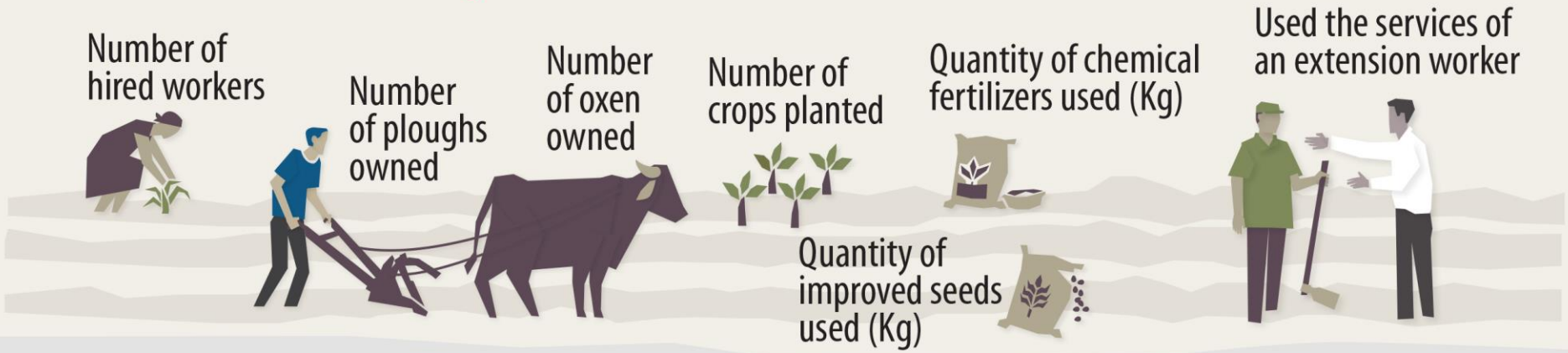
- Many inputs are cross-correlated with each other – can choose one input to represent a cluster of closely-correlated inputs
- Cross-correlations between inputs and outputs are low
- Most predictive inputs have a similar directional effect across outcome variables, yet their impact varies
- Similar results hold across years (3 waves of analysis)



Then, cluster farmers by considering inputs that are highly cross-correlated with outcomes and not with other inputs

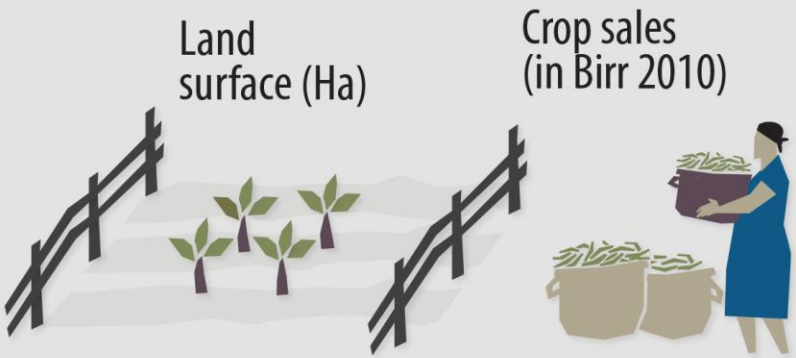
ACTIONABLE VARIABLES

Actionable variables with high cross-correlation across outcomes

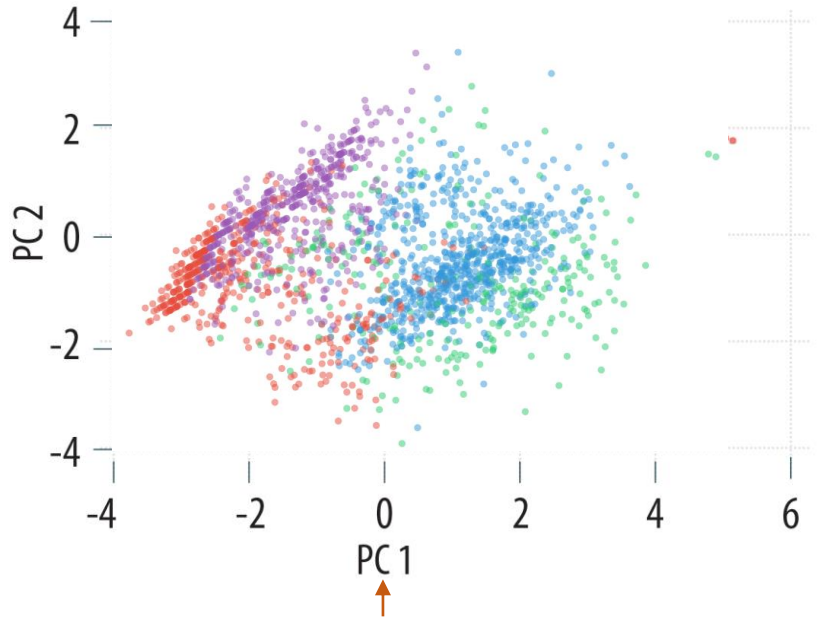


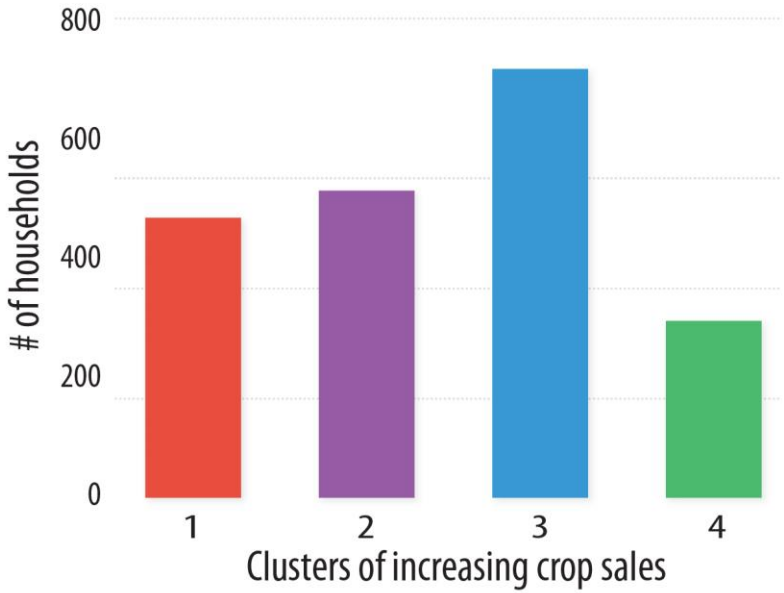
NON-ACTIONABLE VARIABLES

Non-actionable variables with high cross-correlation across outcomes

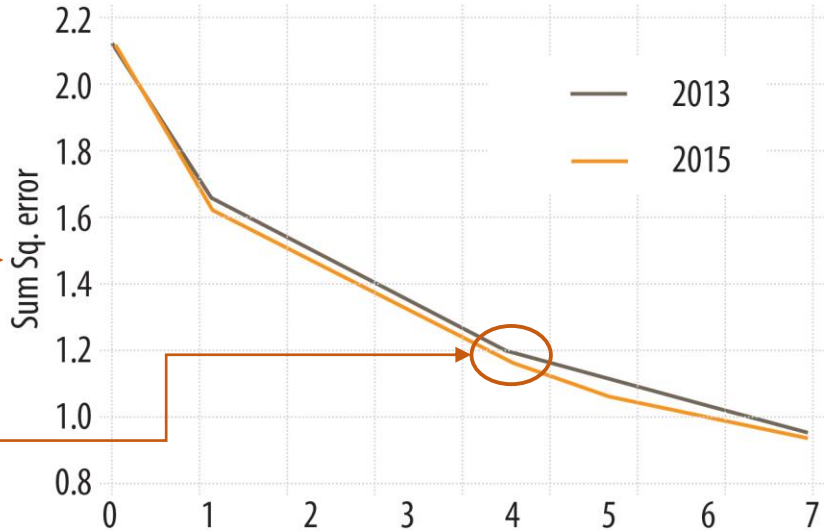


K-means clustering results

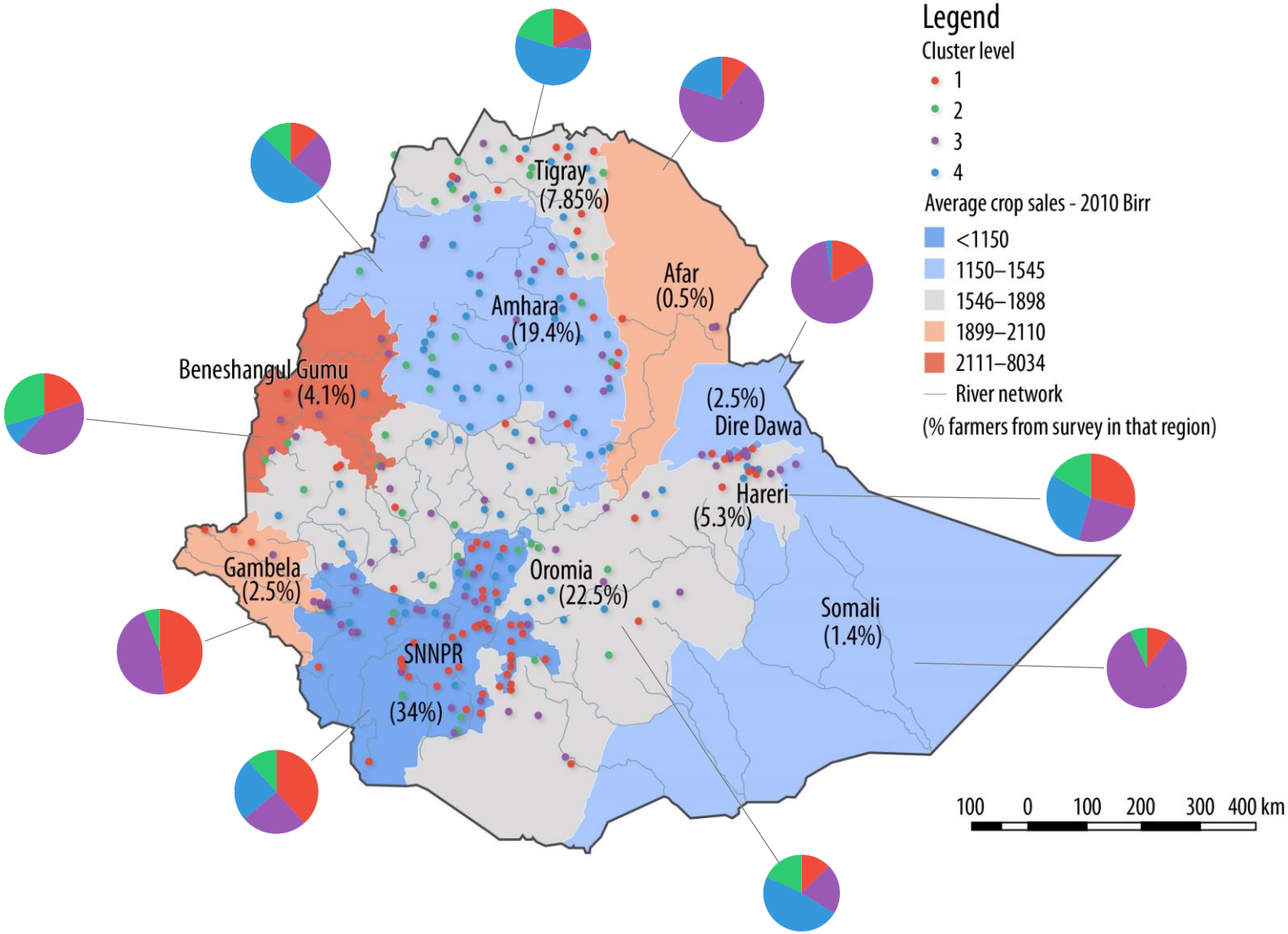




- K-means clustering achieves desired outcome: clusters farmers based on their own unique set of actionable variables most correlated with outcomes and not with other input variable
- Clustering consistent over time
- We pick: number of clusters = **4**

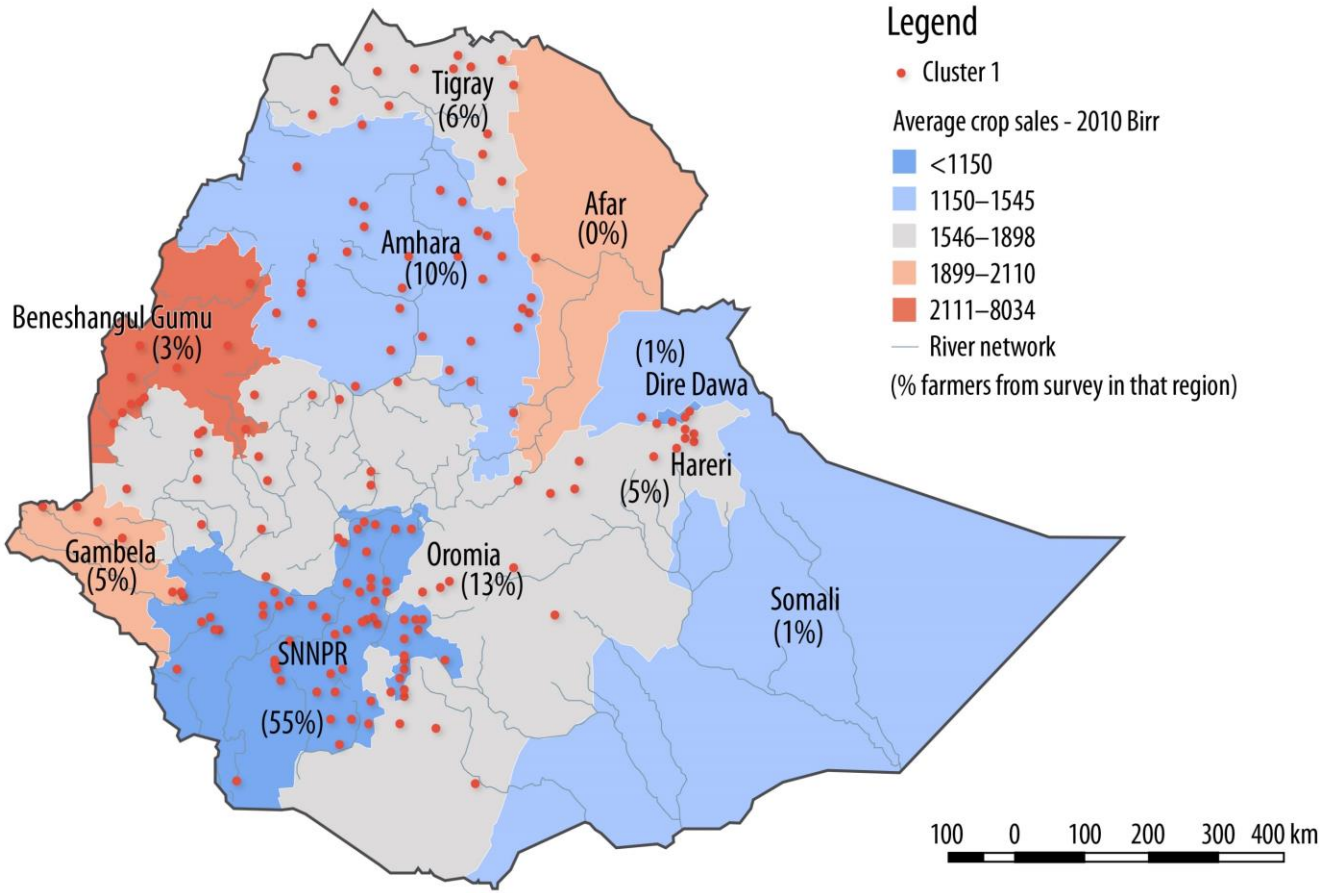


Where are the clusters?



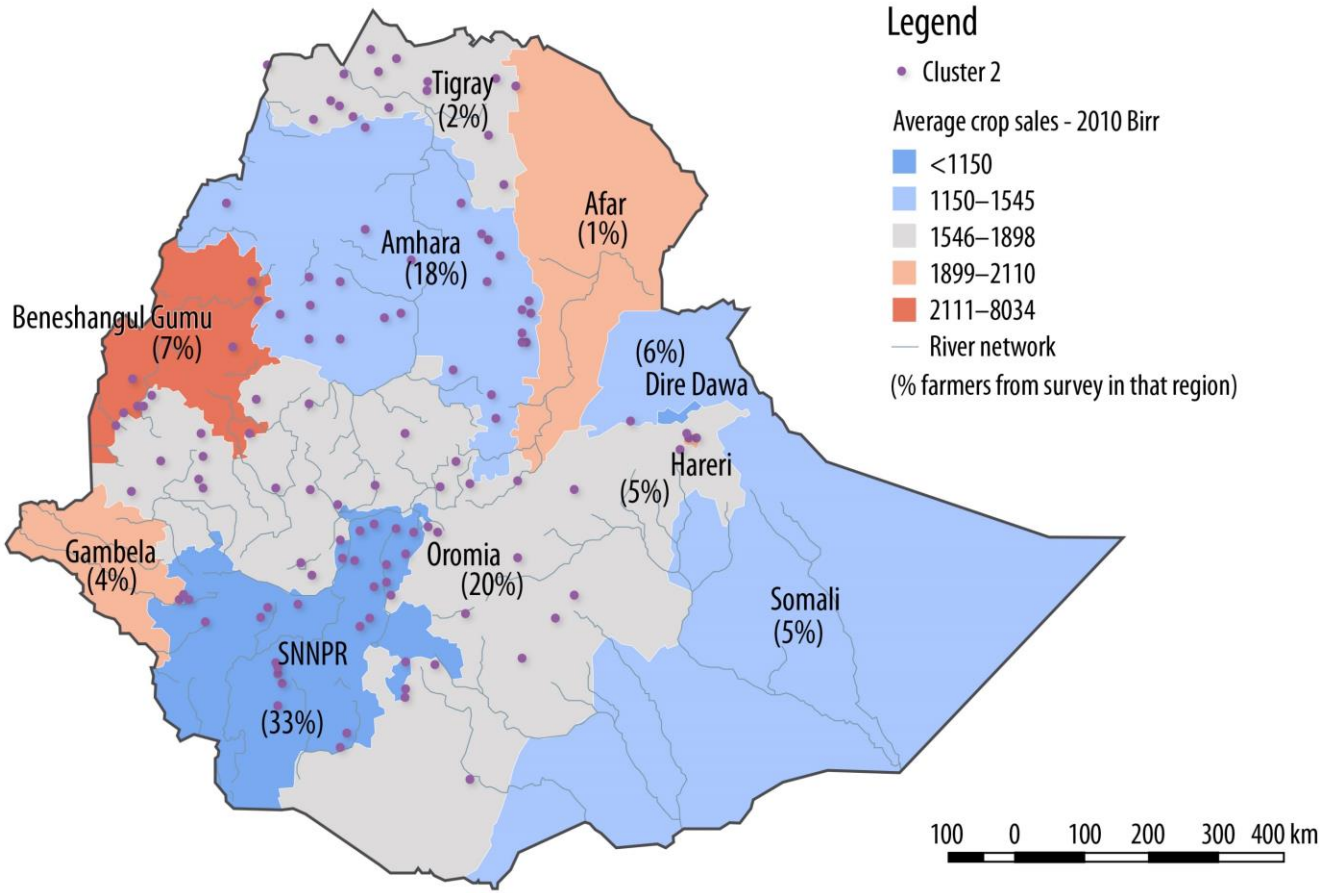
WORLD BANK GROUP

Cluster 1



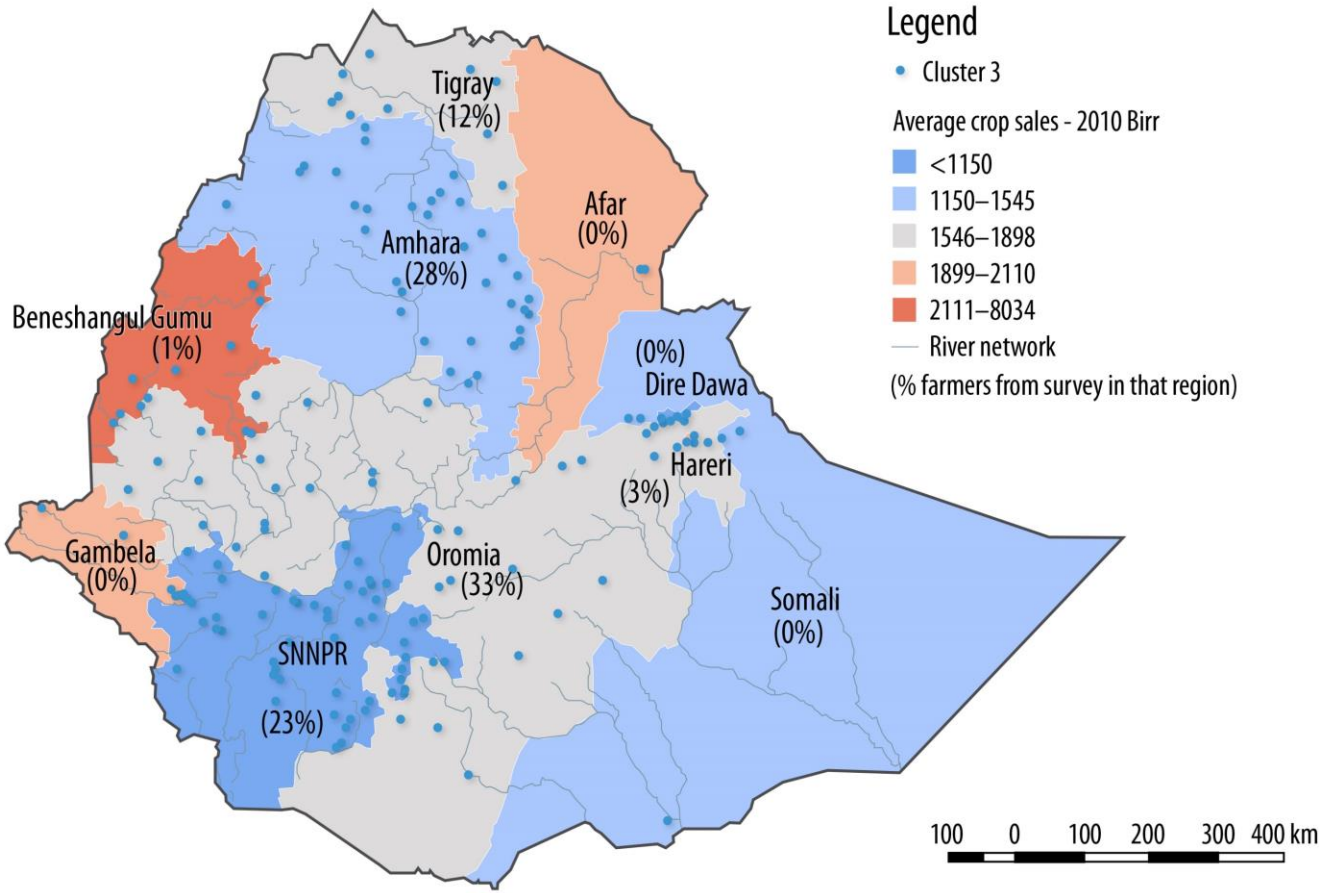
WORLD BANK GROUP

Cluster 2



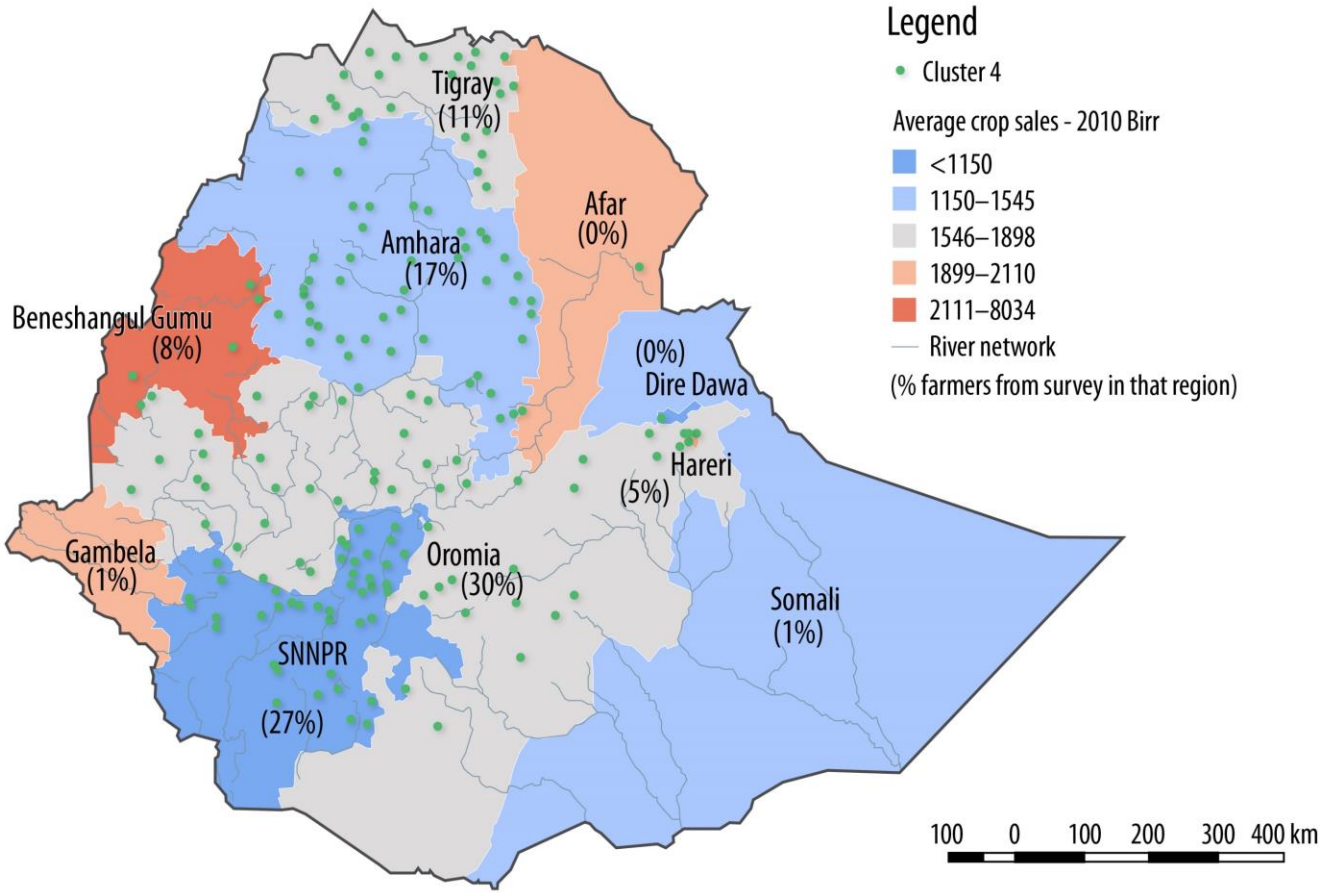
WORLD BANK GROUP

Cluster 3



WORLD BANK GROUP

Cluster 4



WORLD BANK GROUP

How are the clusters different? *(actionable variables)*

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Number of hired workers	0.32	0.17	0.18	17.69
Number of oxen owned	0.16	0.96	1.76	2.06
Number of plough owned	0.32	0.78	1.22	1.24
Number of crops planted	3.64	2.95	2.84	2.78
Quantity of chemical fertilizers used (in kg)	22.99	7.45	378.15	343.63
Quantity of improved seeds used (in kg)	2.12	0.93	11.83	12.77
% of households who own land certificate	0.43	0.54	0.67	0.62
% of households using extension programs	0.25	0.06	0.80	0.41
% of households who saved	0.12	0.15	0.15	0.21
% of households using credit services	0.12	0.13	0.28	0.26

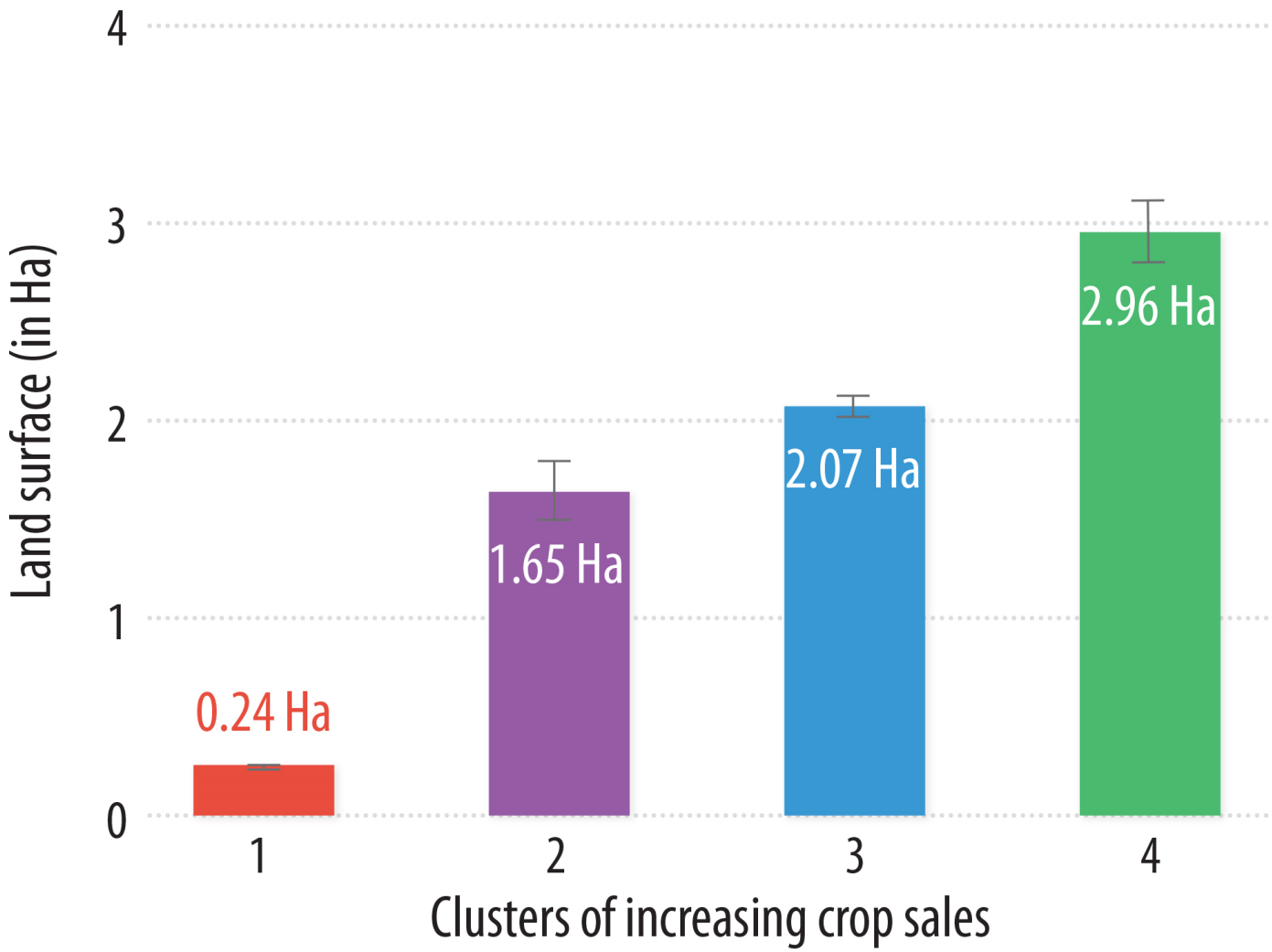
How are the clusters different? (non-actionable variables and outcomes)

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Land surface (in Ha)	0.24	1.65	2.07	2.96
Crop sales (in Birr 2010)	710.99	1279.19	1524.49	2427.23
Non-food expenditure (in Birr 2010)	1064.44	1236.68	1773.64	2397.29
% households using medical assistance	0.2	0.22	0.23	0.28
% households without food deficiencies	0.47	0.7	0.78	0.84

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

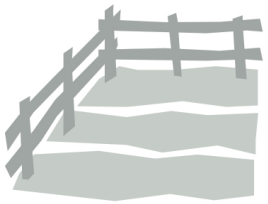
First, variation in non-actionable variables

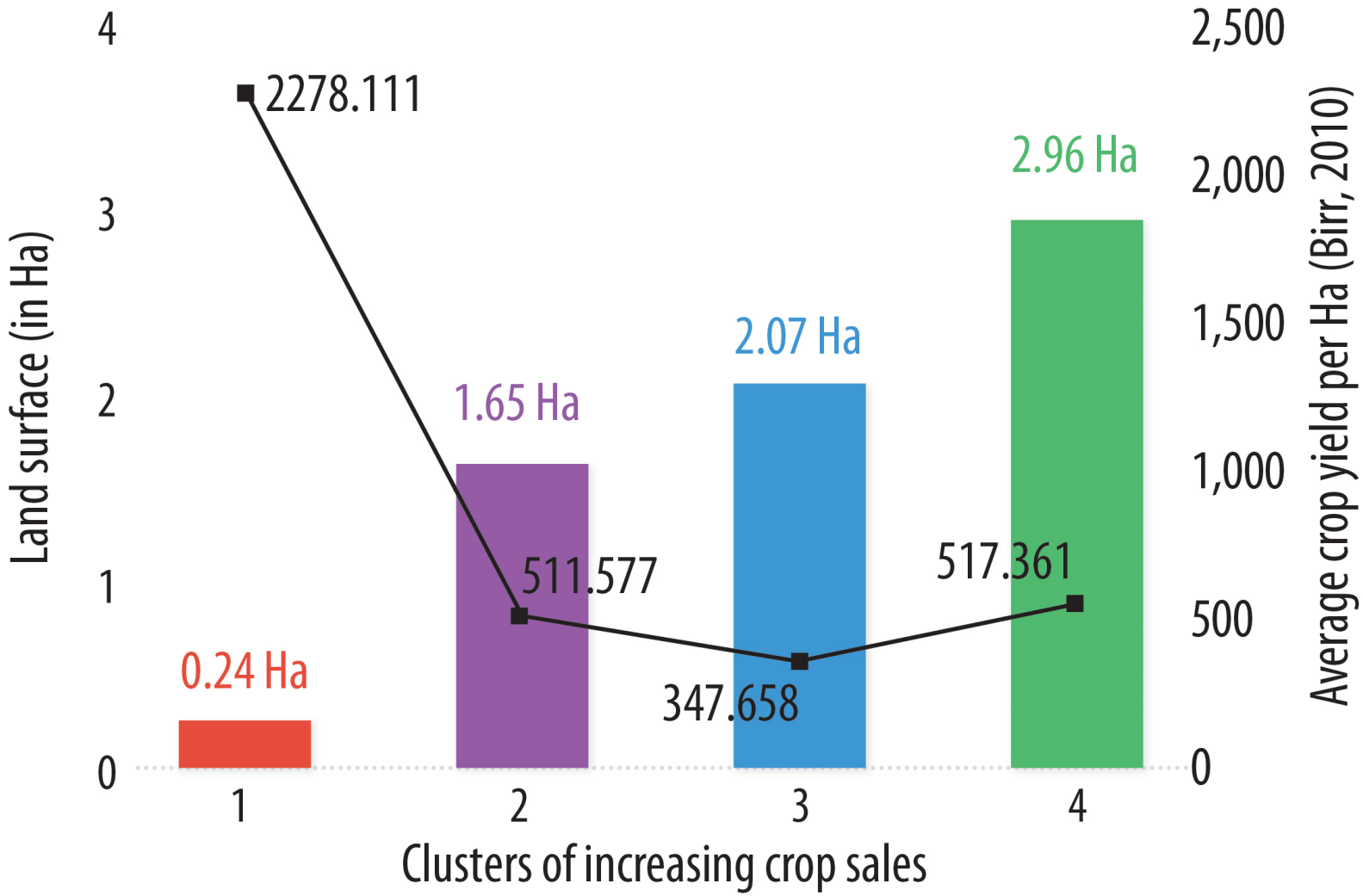
Variation in land surface across clusters



WORLD BANK GROUP

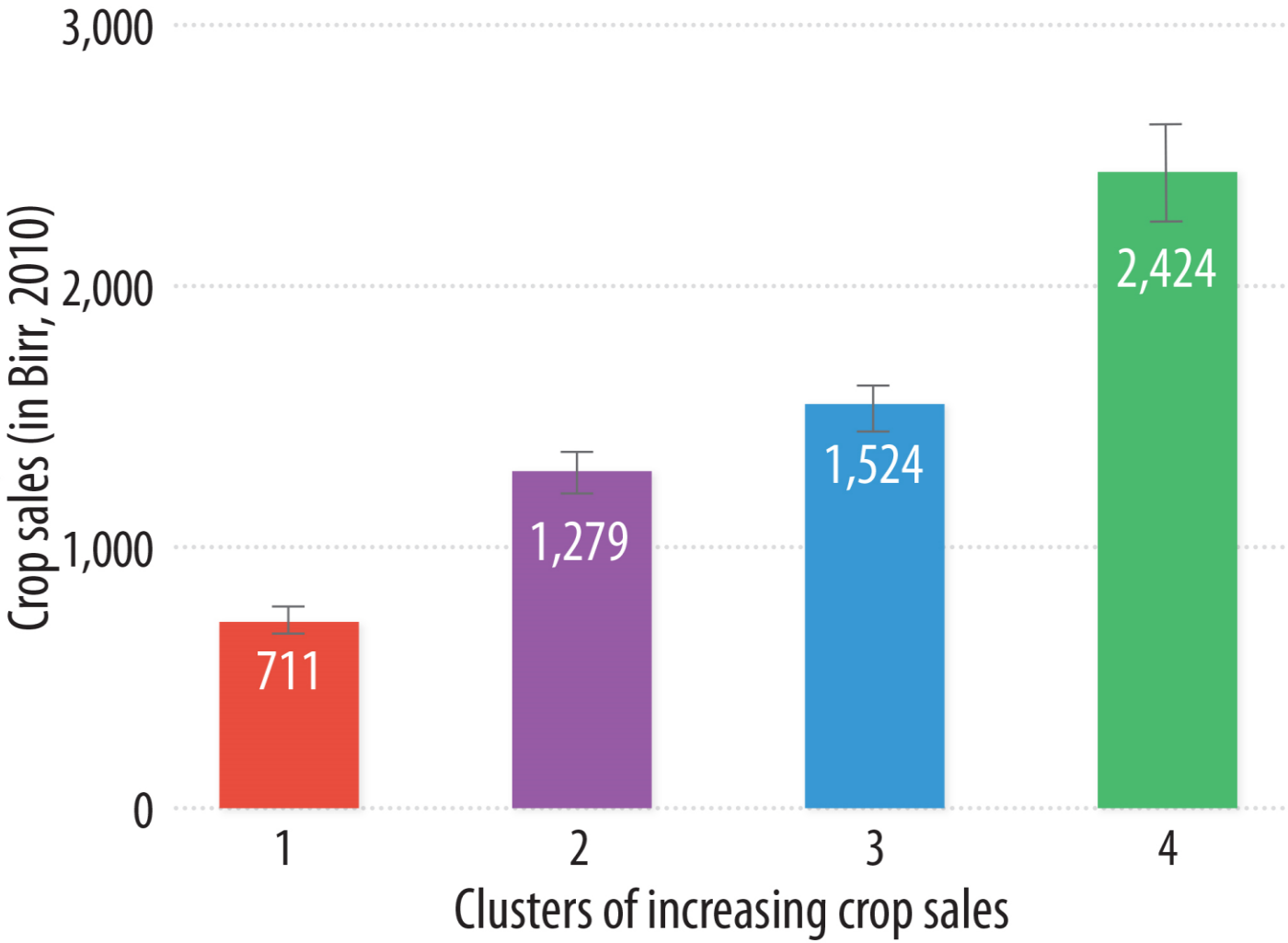
Variation in land surface across clusters





WORLD BANK GROUP

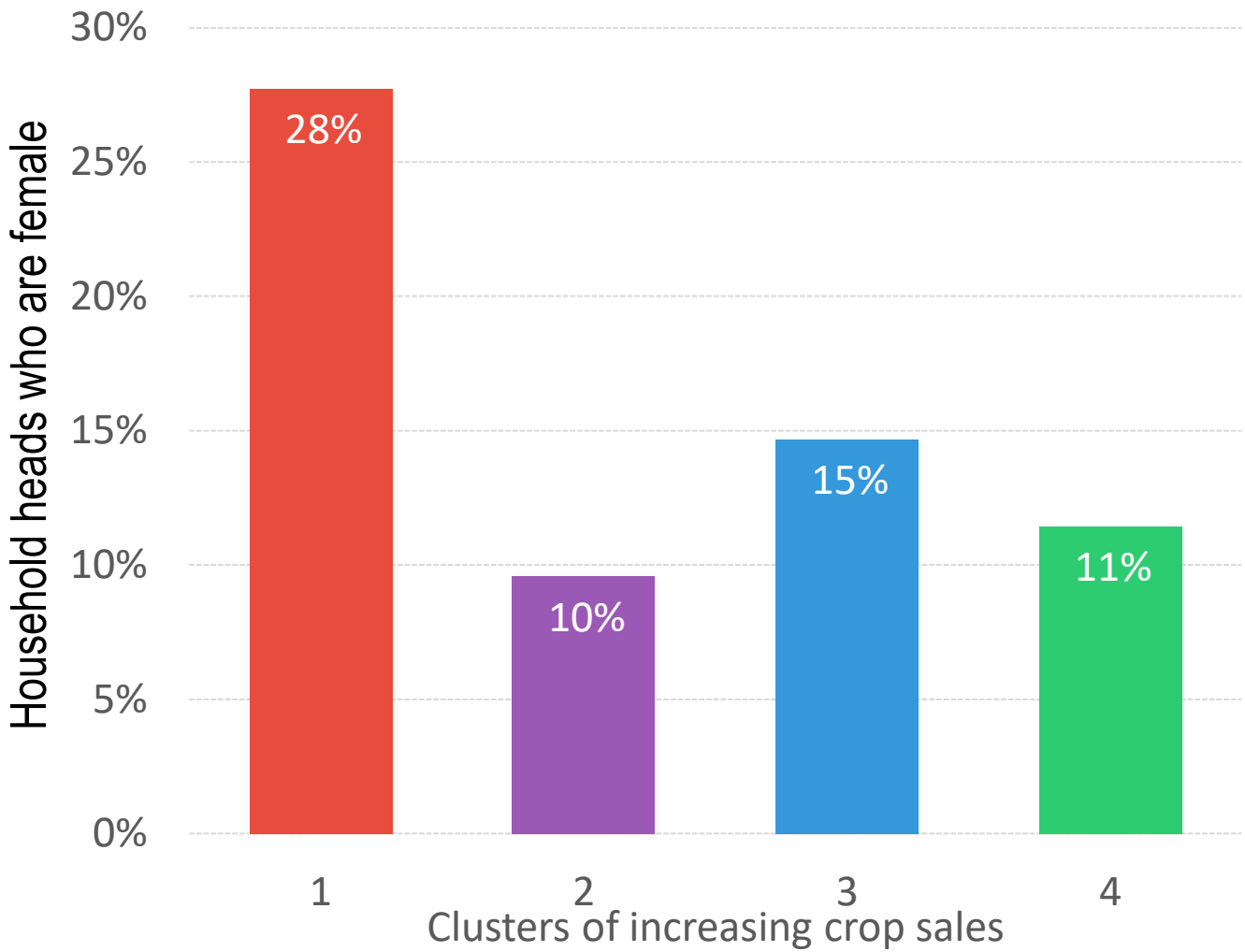
Variation in crop sales (in Birr 2010) across clusters



Crop sales and productivity are **72%** correlated with each other so can analyze them together

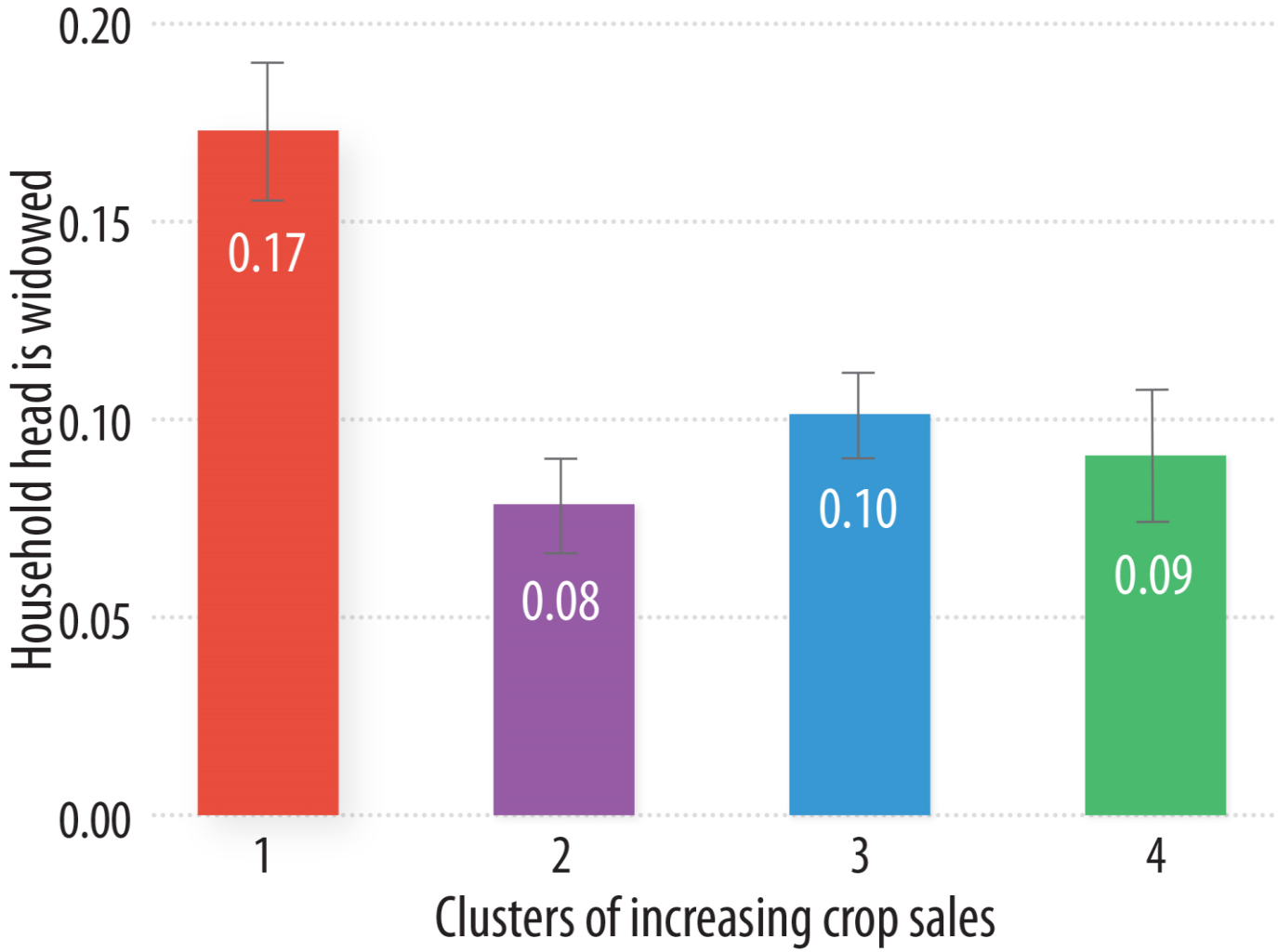
WORLD BANK GROUP

Variation in % households heads who are female across clusters



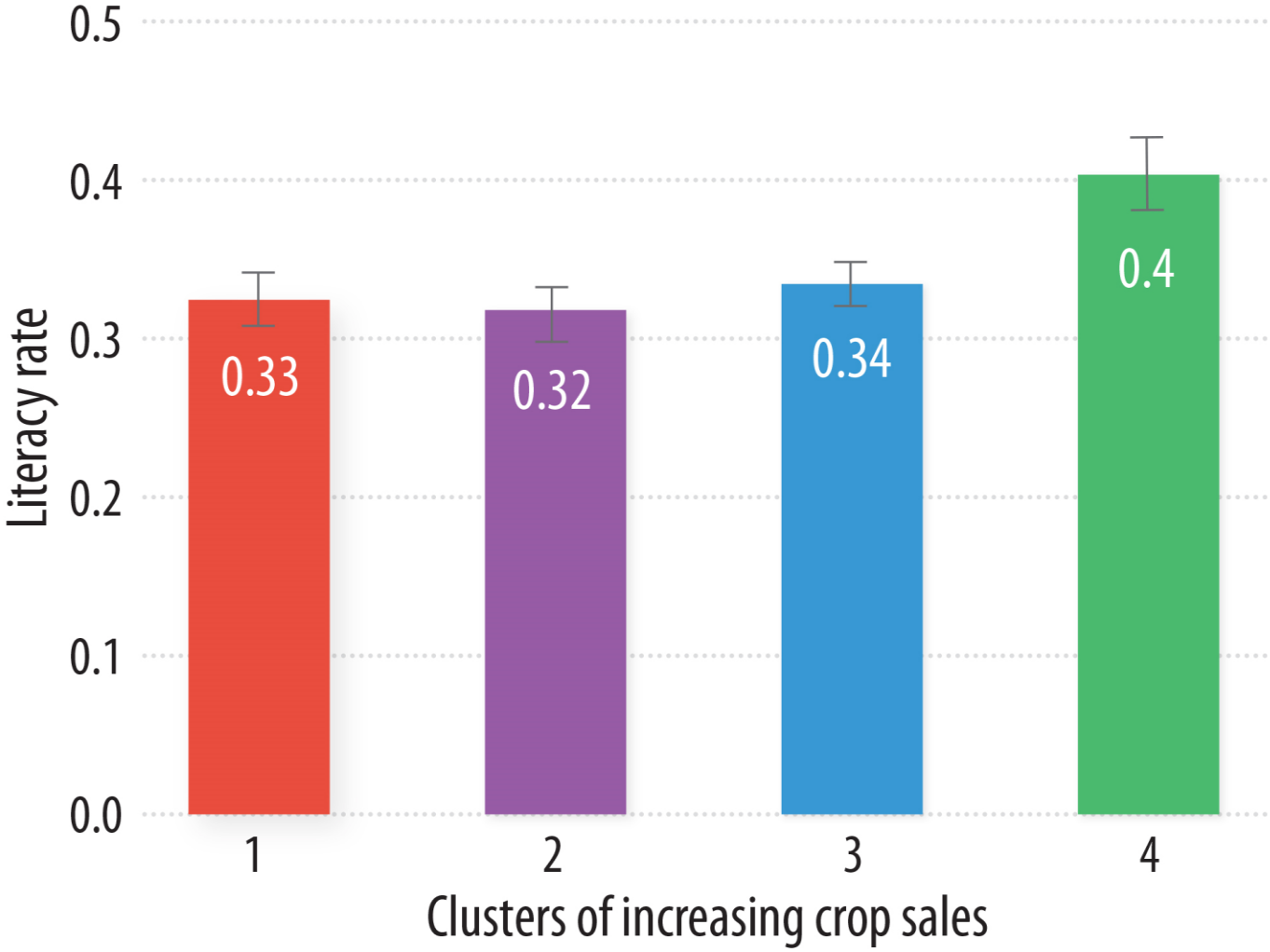
WORLD BANK GROUP

Variation in % widowed households across clusters



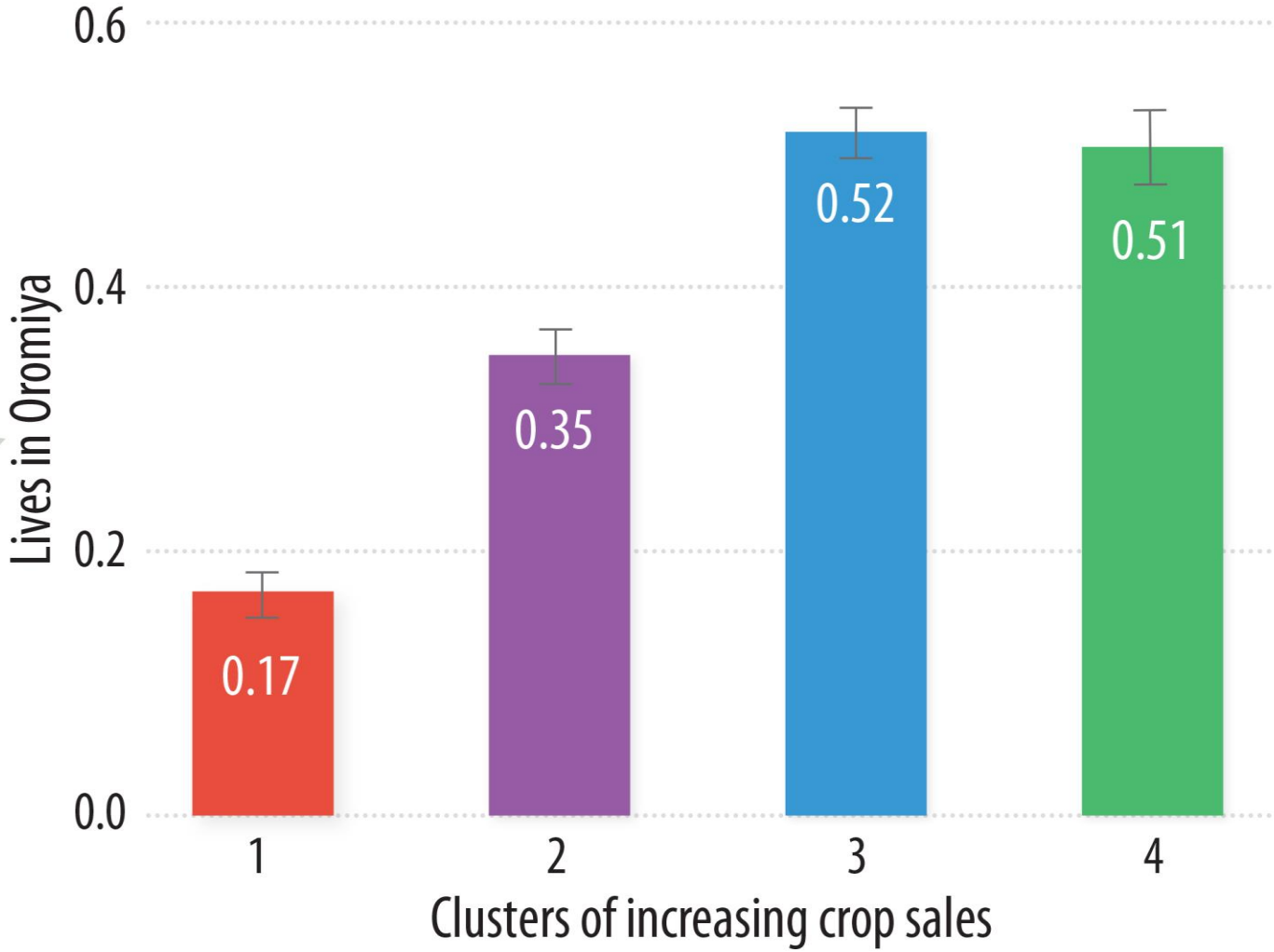
WORLD BANK GROUP

Variation in household literacy rate across clusters



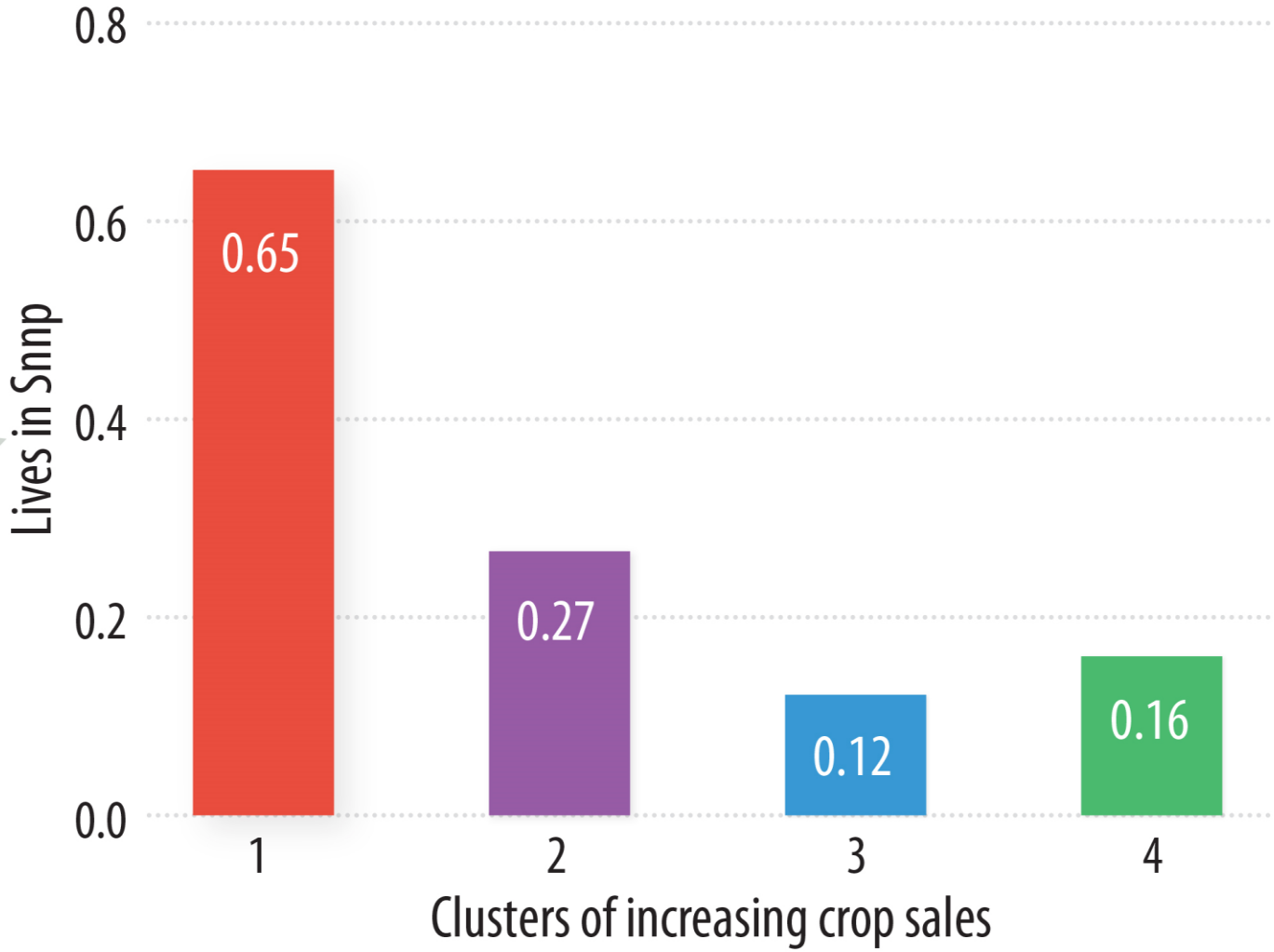
WORLD BANK GROUP

Proportion of households across clusters who lives in Oromiya



WORLD BANK GROUP

Proportion of households across clusters who lives in SNNPR

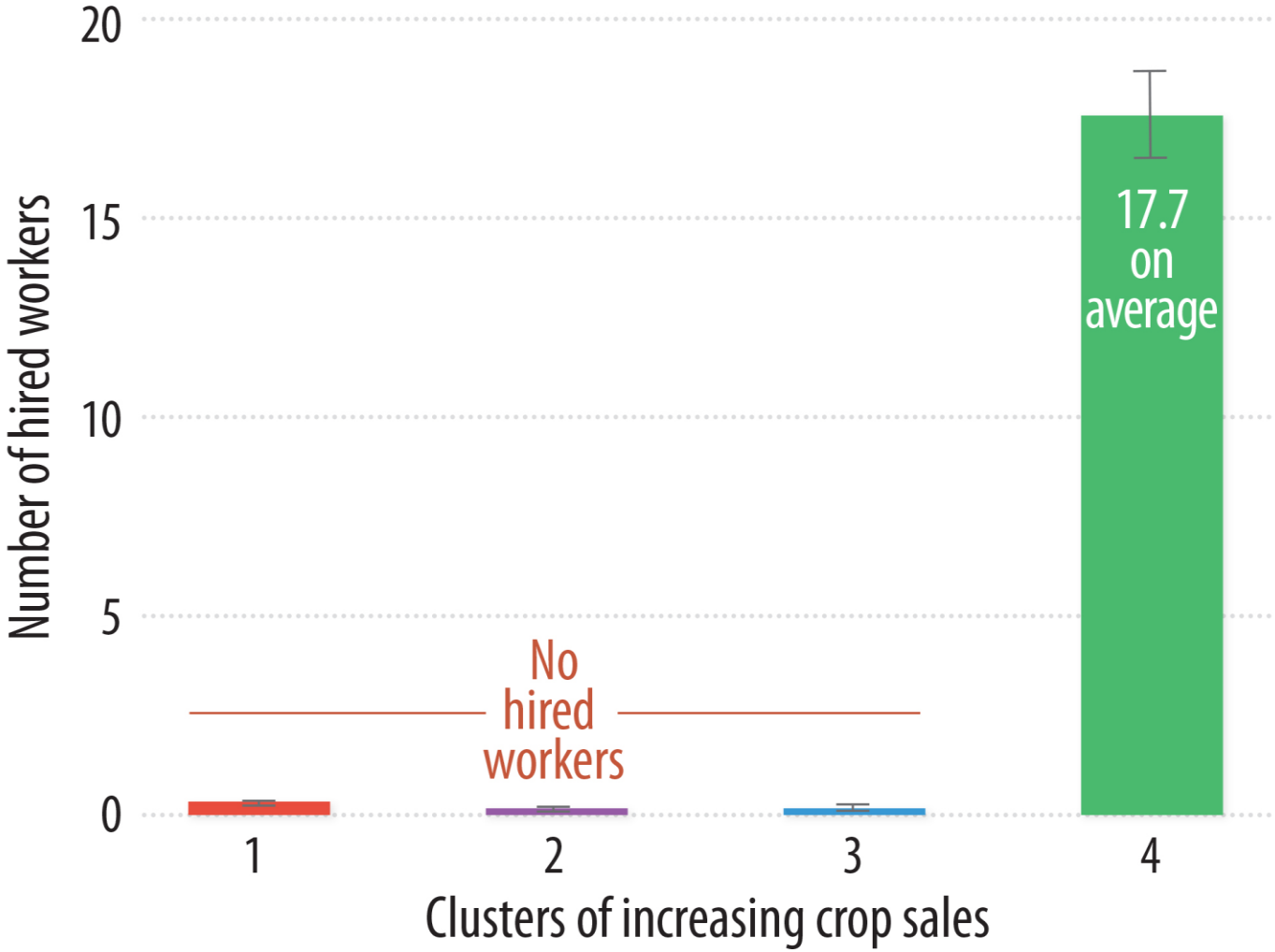


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

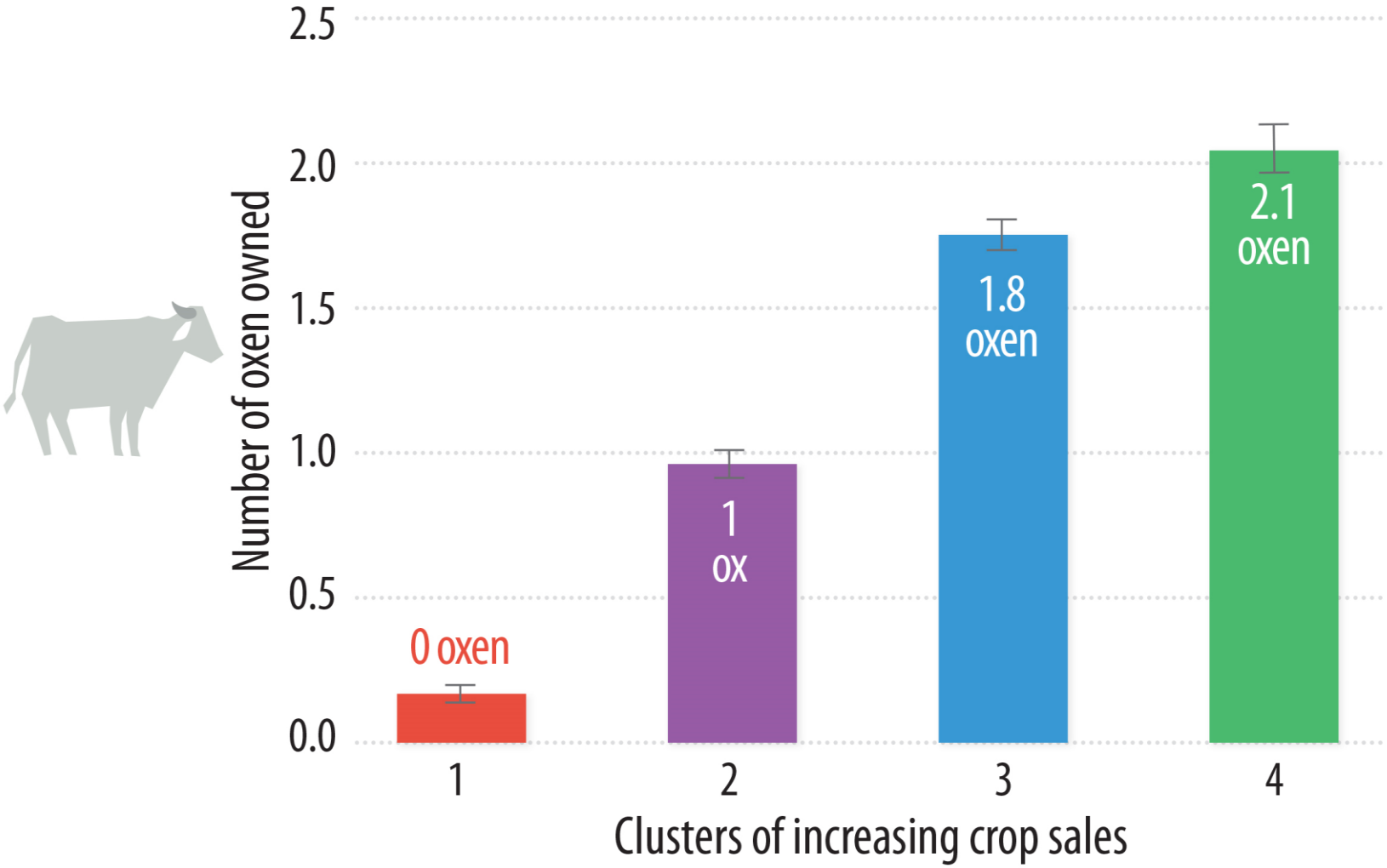
Second, variation in actionable variables

Variation in hiring workers across clusters



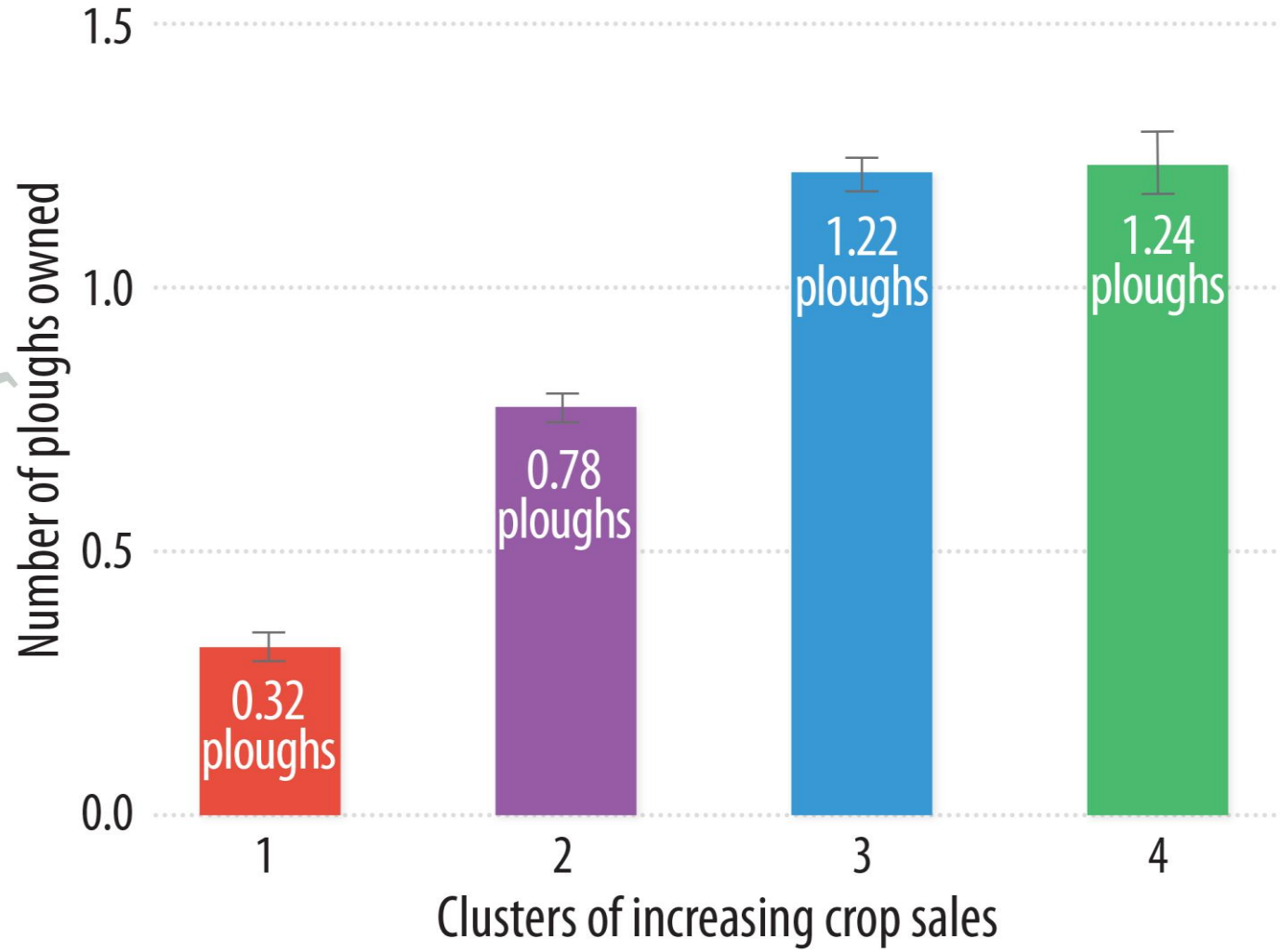
WORLD BANK GROUP

Variation in agricultural equipment (oxen) across clusters



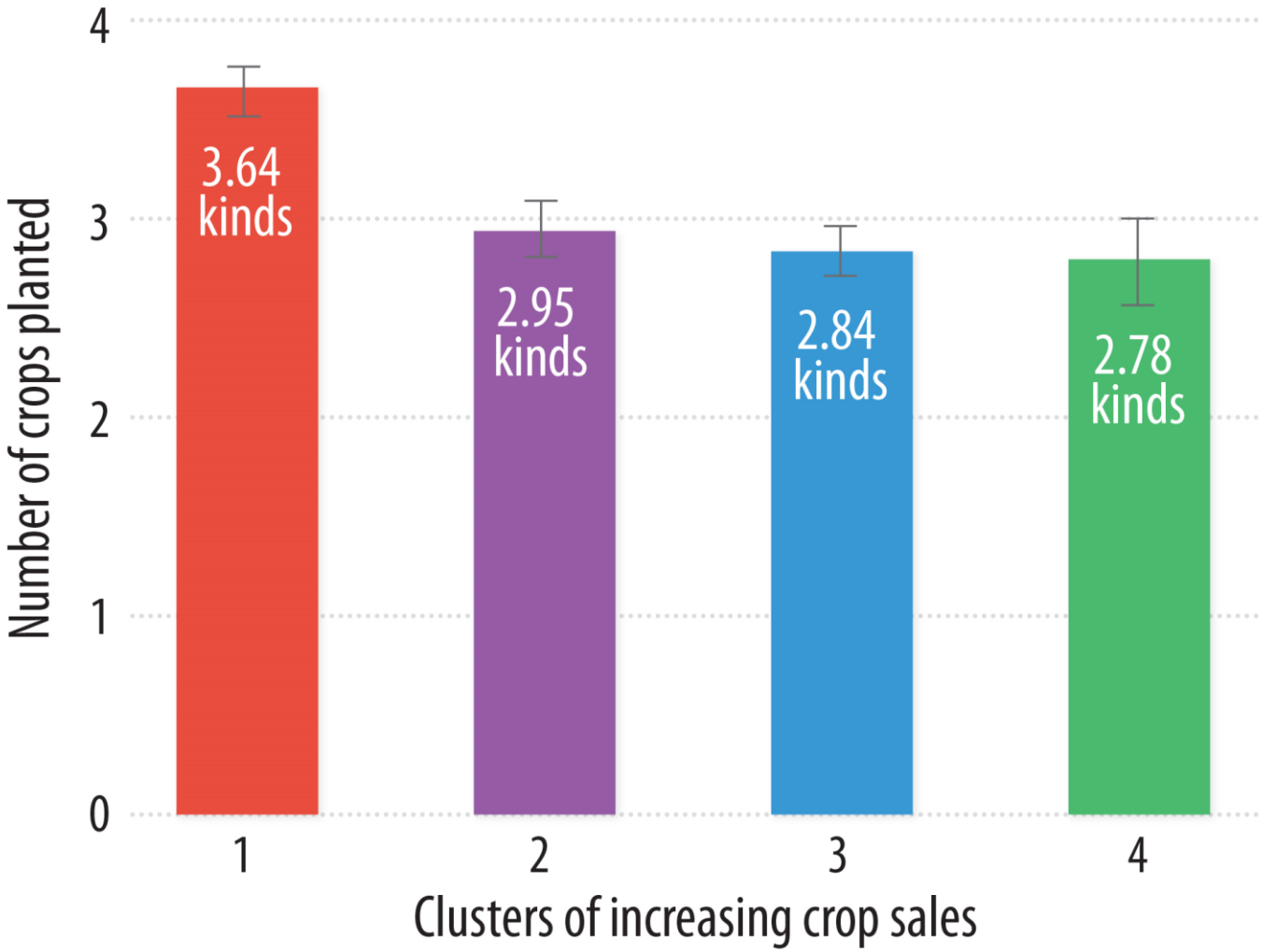
WORLD BANK GROUP

The variation in agricultural equipment (ploughs) across clusters



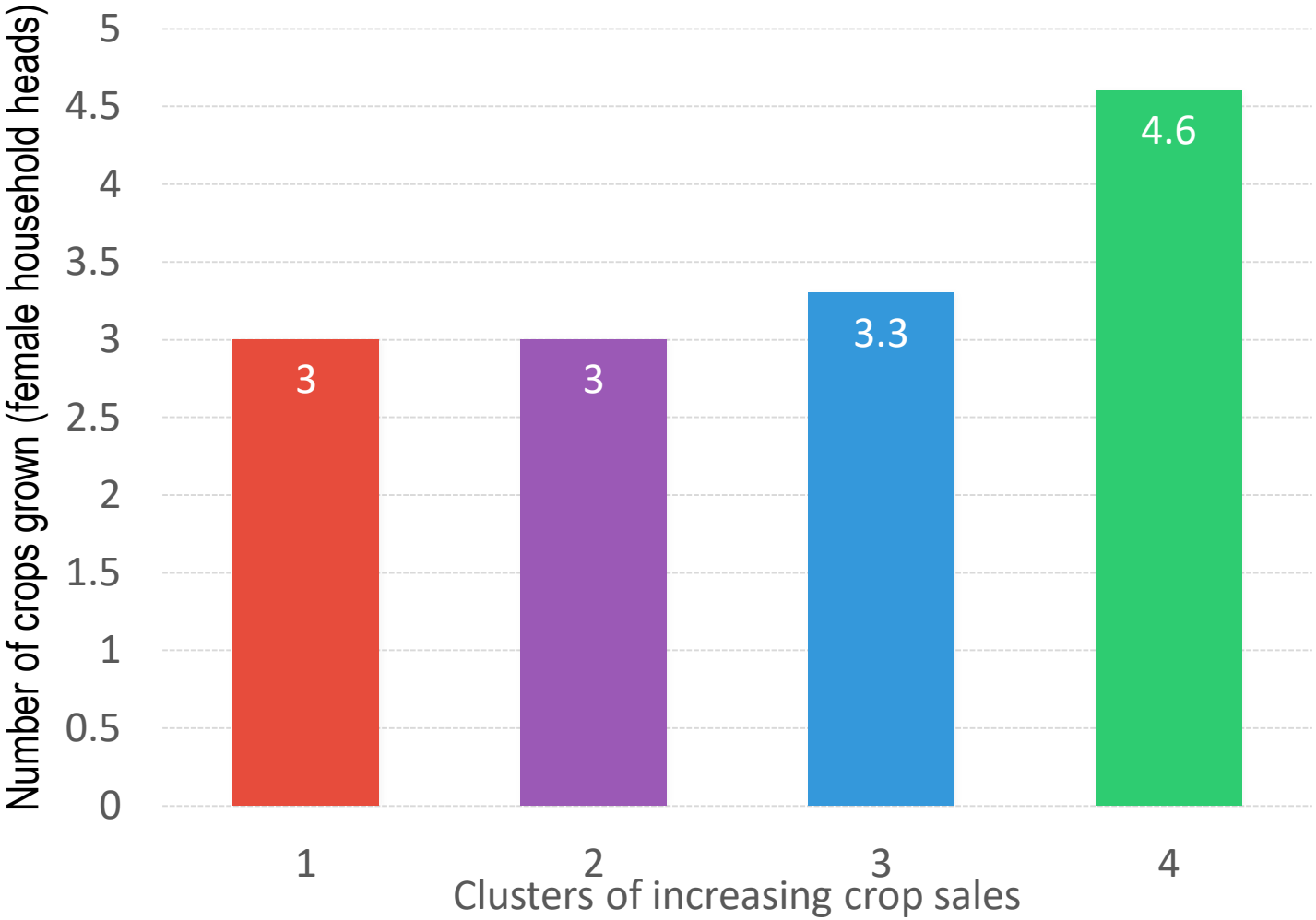
WORLD BANK GROUP

The variation in number of crops planted (male-headed household) across clusters



WORLD BANK GROUP

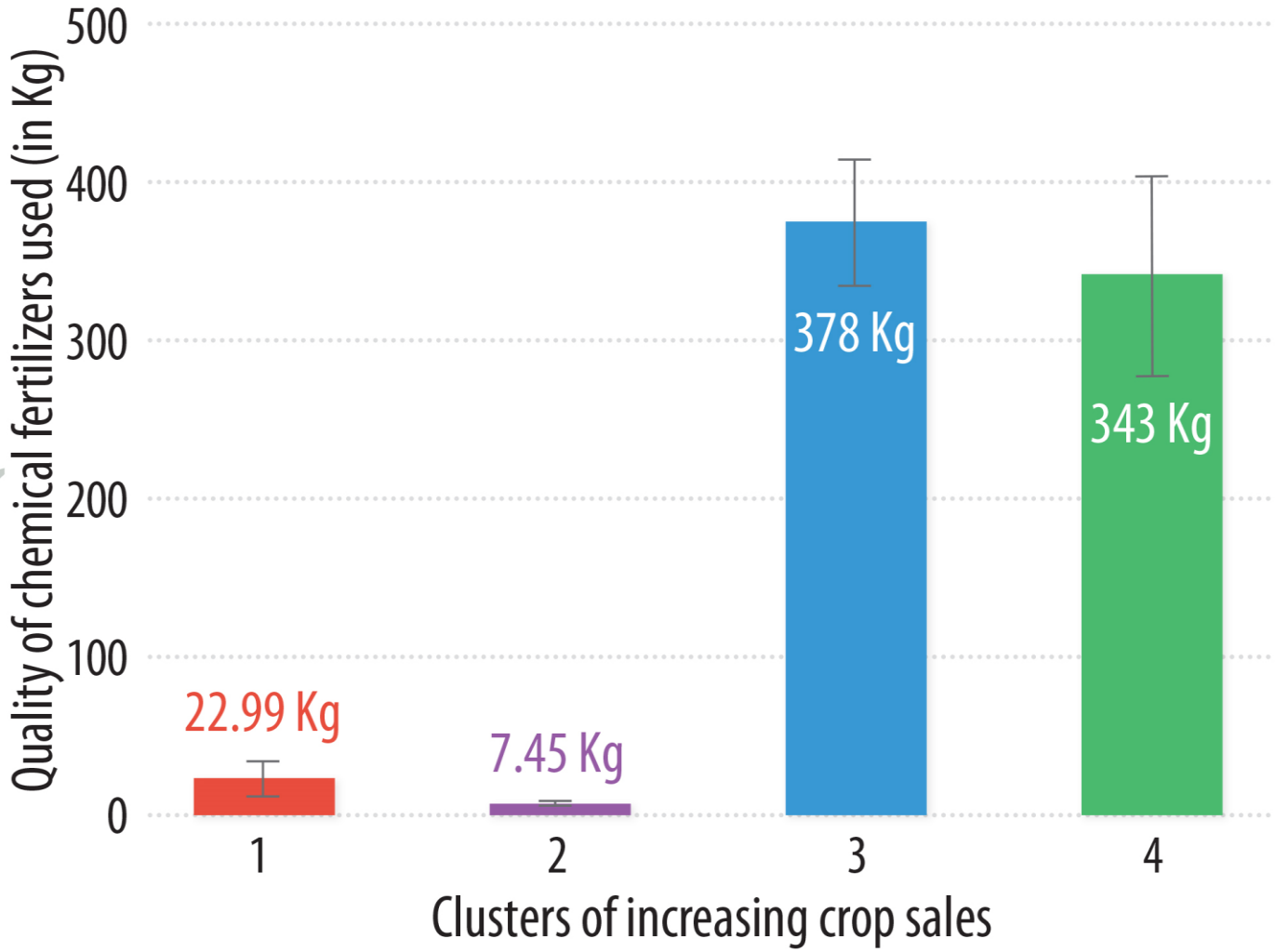
The variation in number of crops planted (female-headed household) across clusters



Female-headed households show a different pattern that warrants further investigation

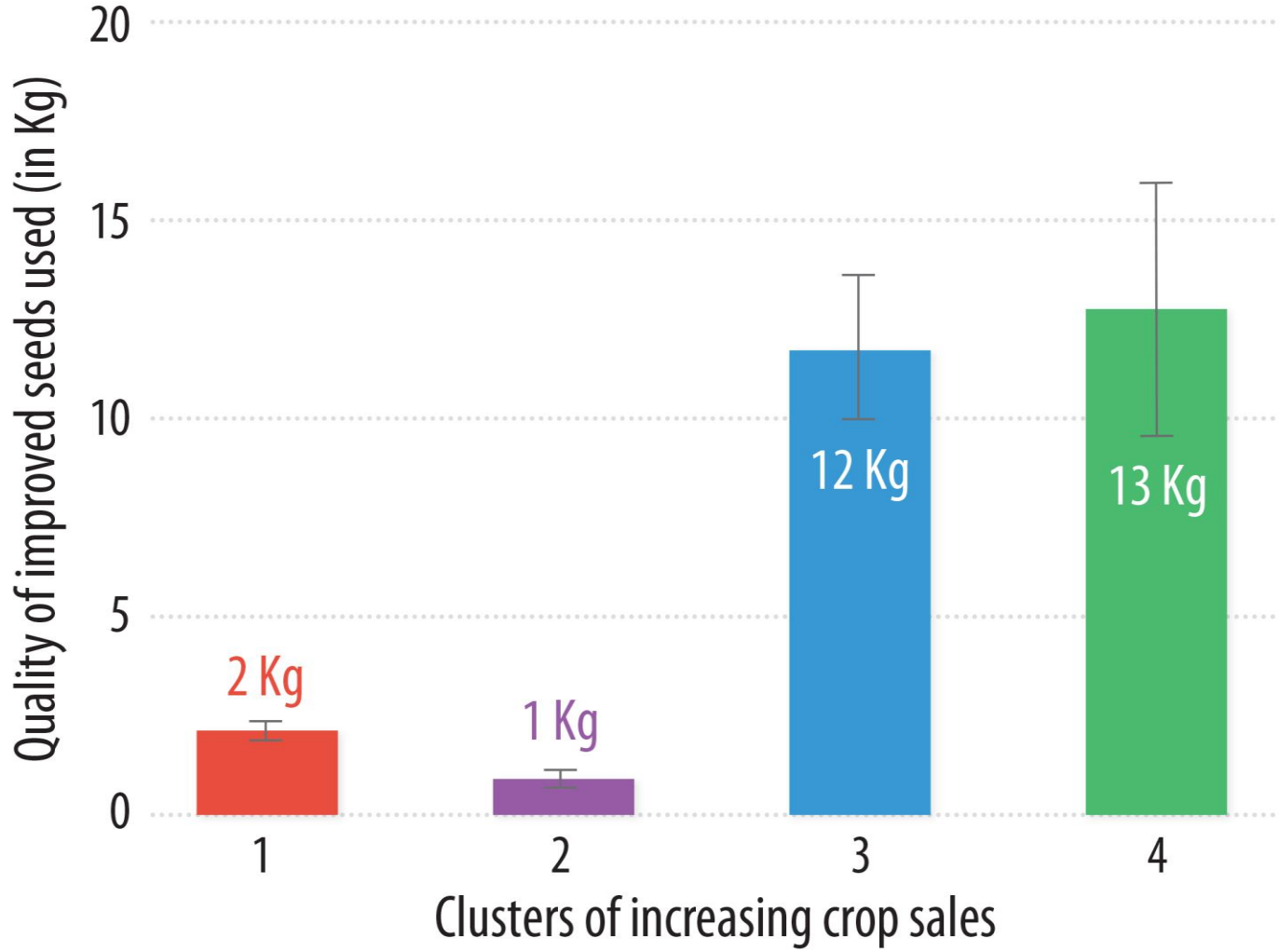
WORLD BANK GROUP

The variation in the use of chemical fertilizers across clusters



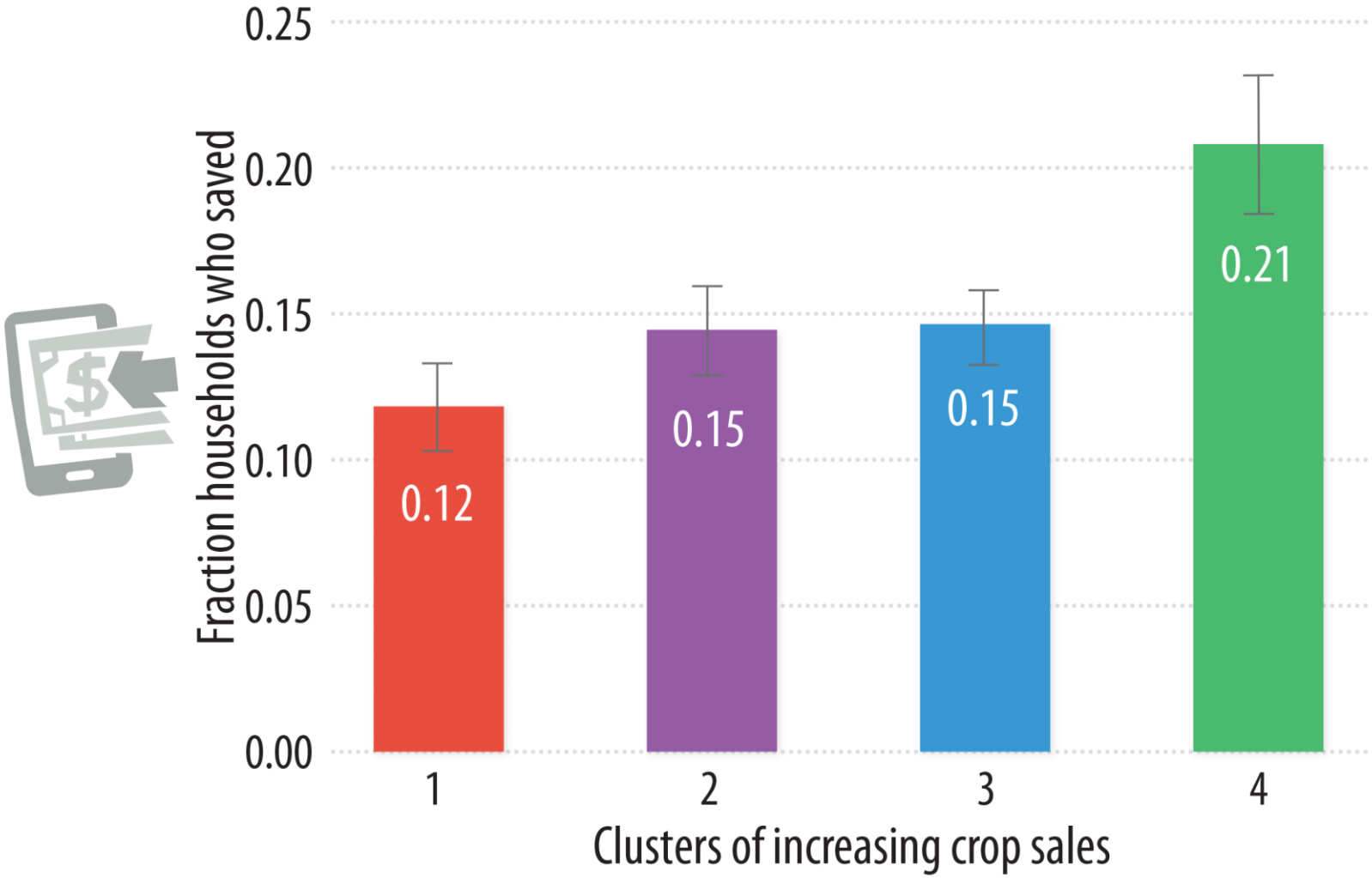
WORLD BANK GROUP

The variation in quantity of improved seeds used across clusters



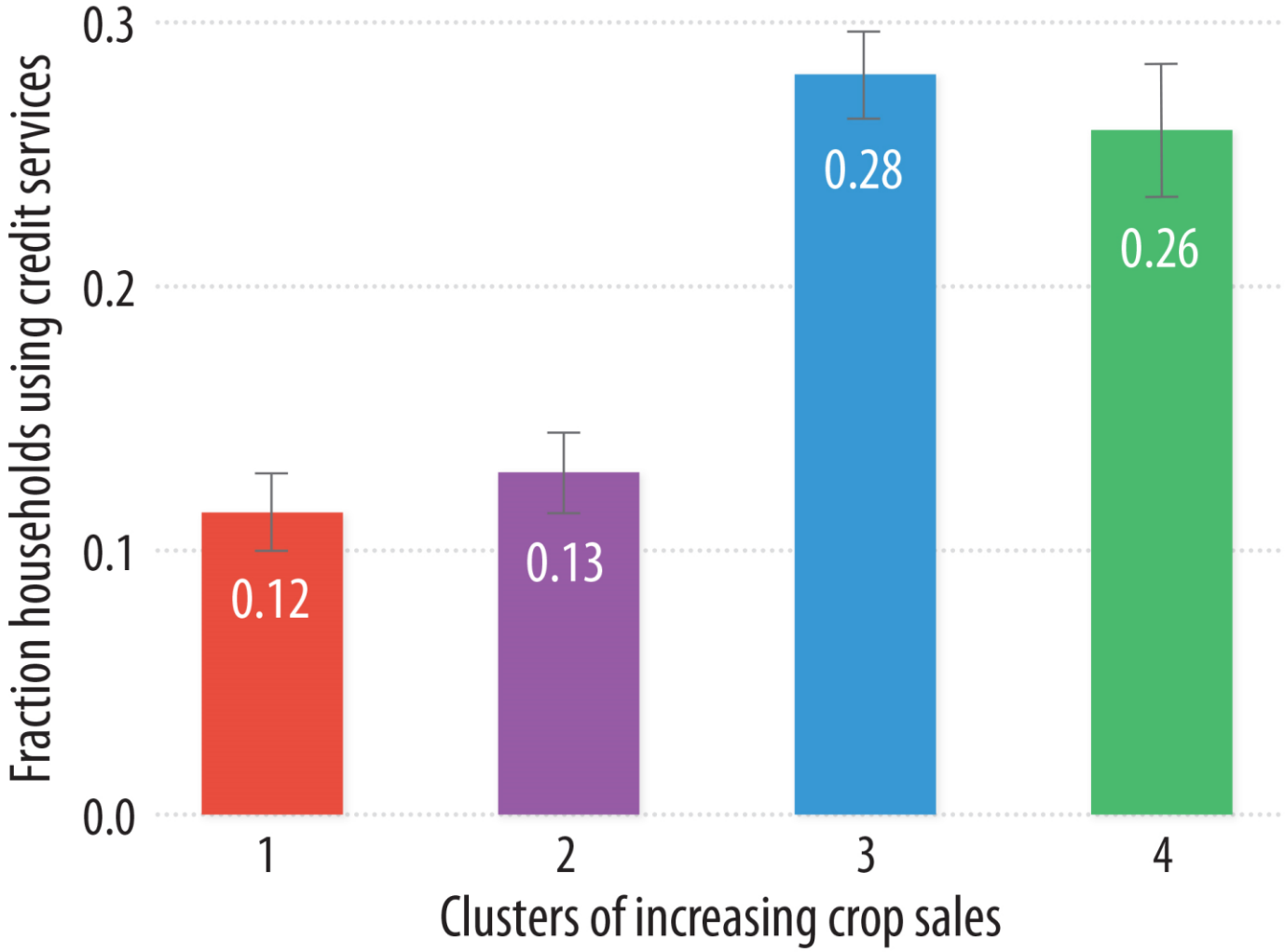
WORLD BANK GROUP

The variation in % of households who saved across clusters



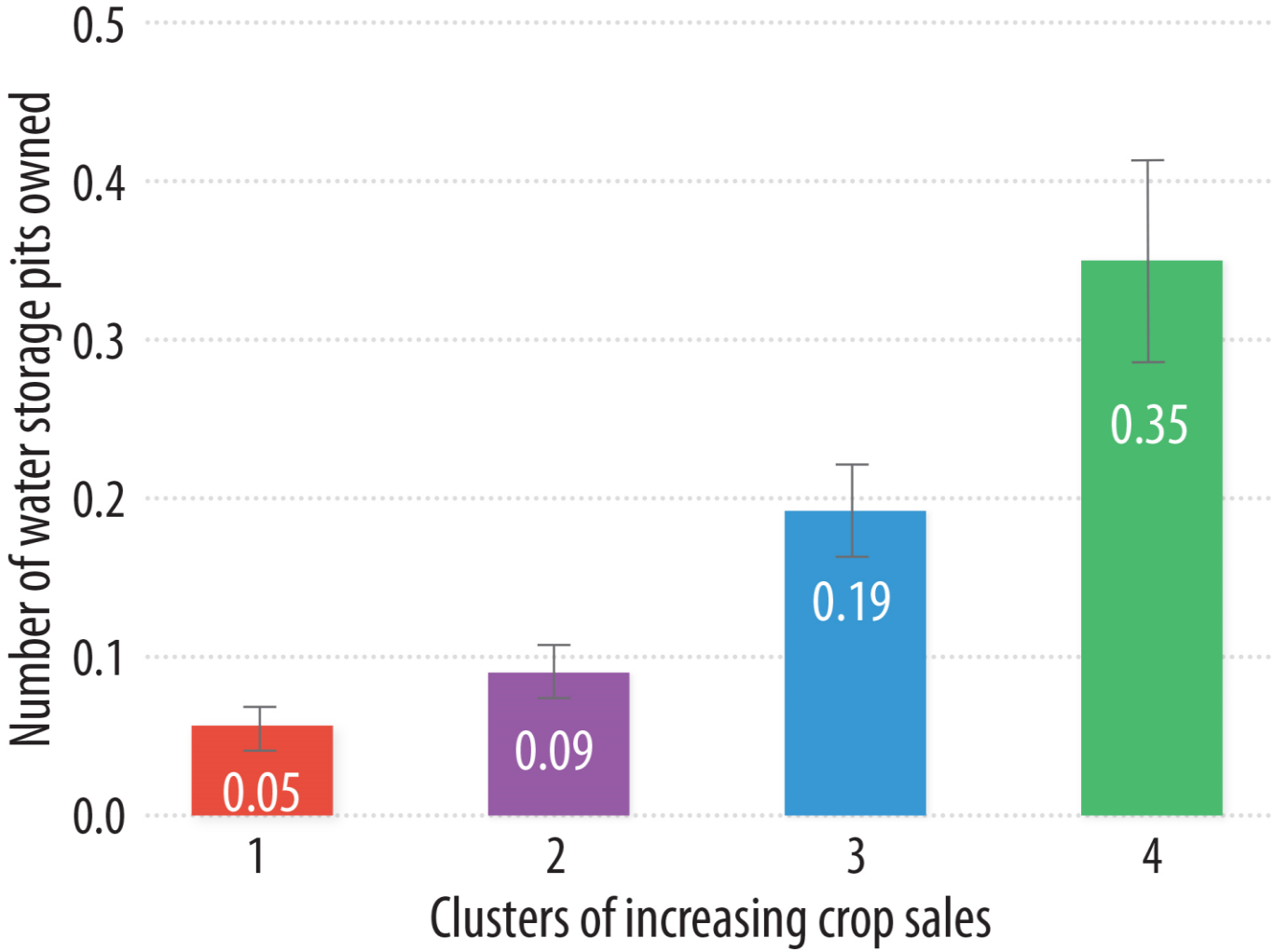
WORLD BANK GROUP

The variation in credit used across clusters



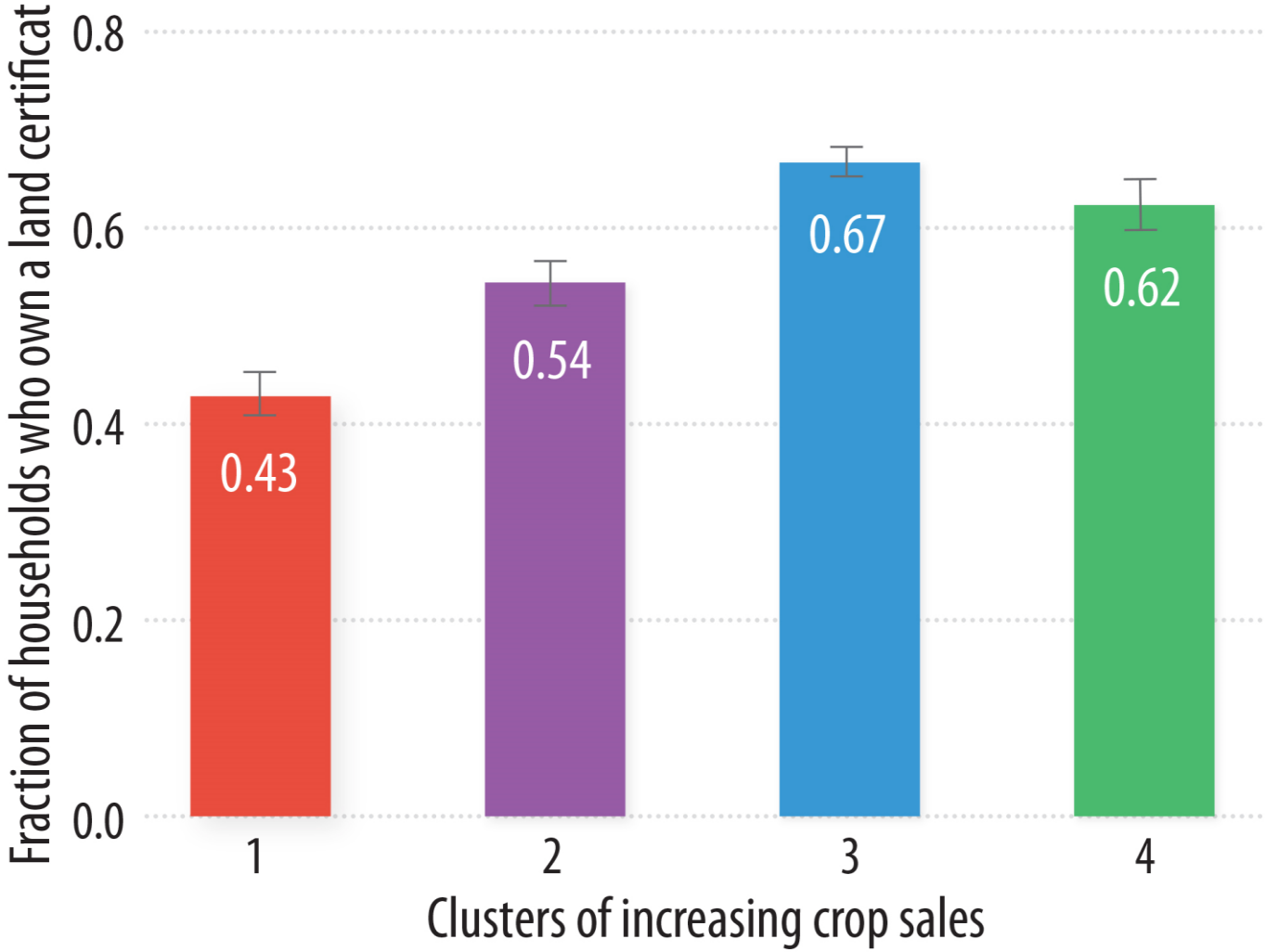
WORLD BANK GROUP

The variation number of water storage pits owned across clusters



WORLD BANK GROUP

The variation % of households who own a land certificate across clusters

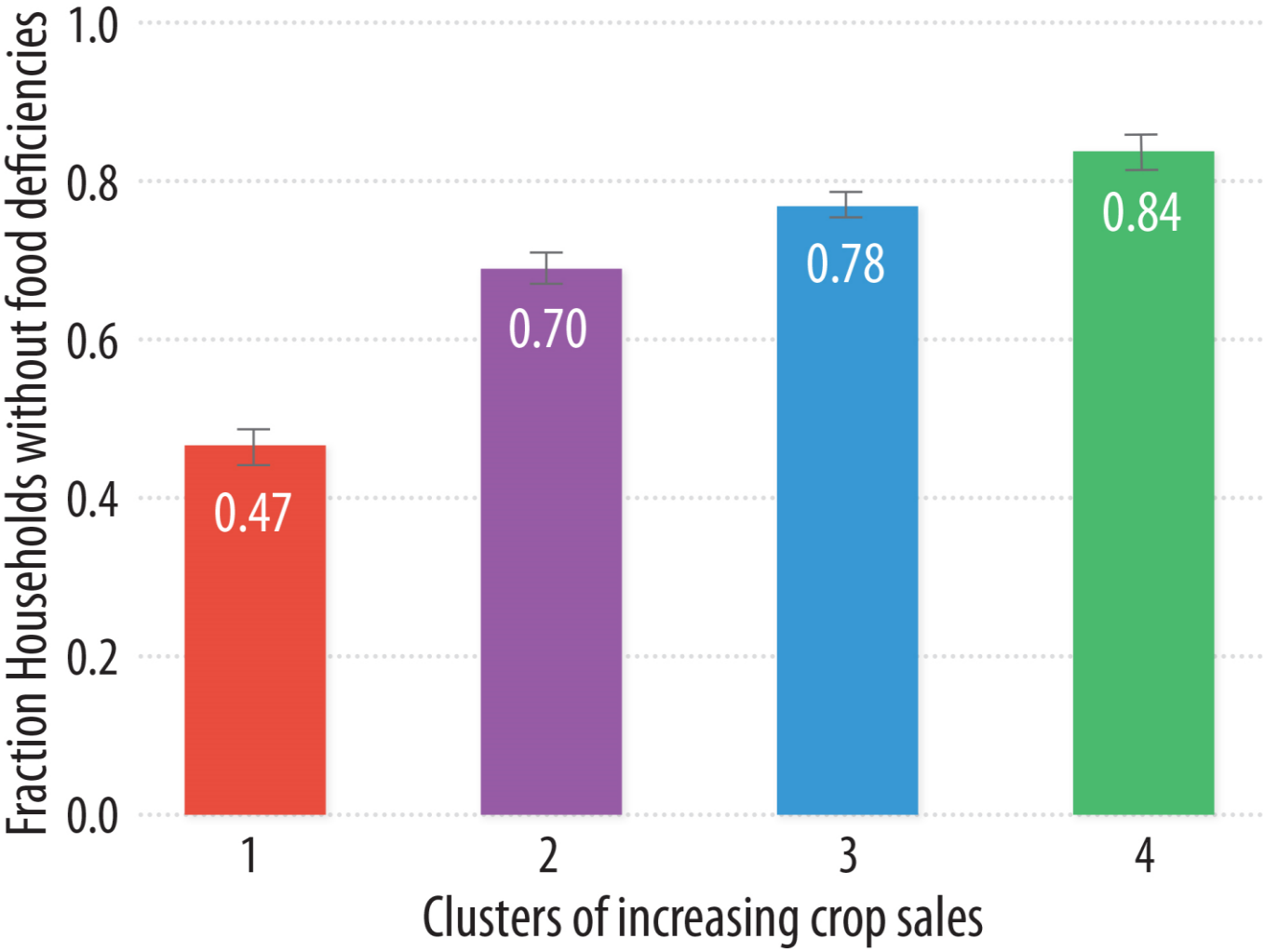


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

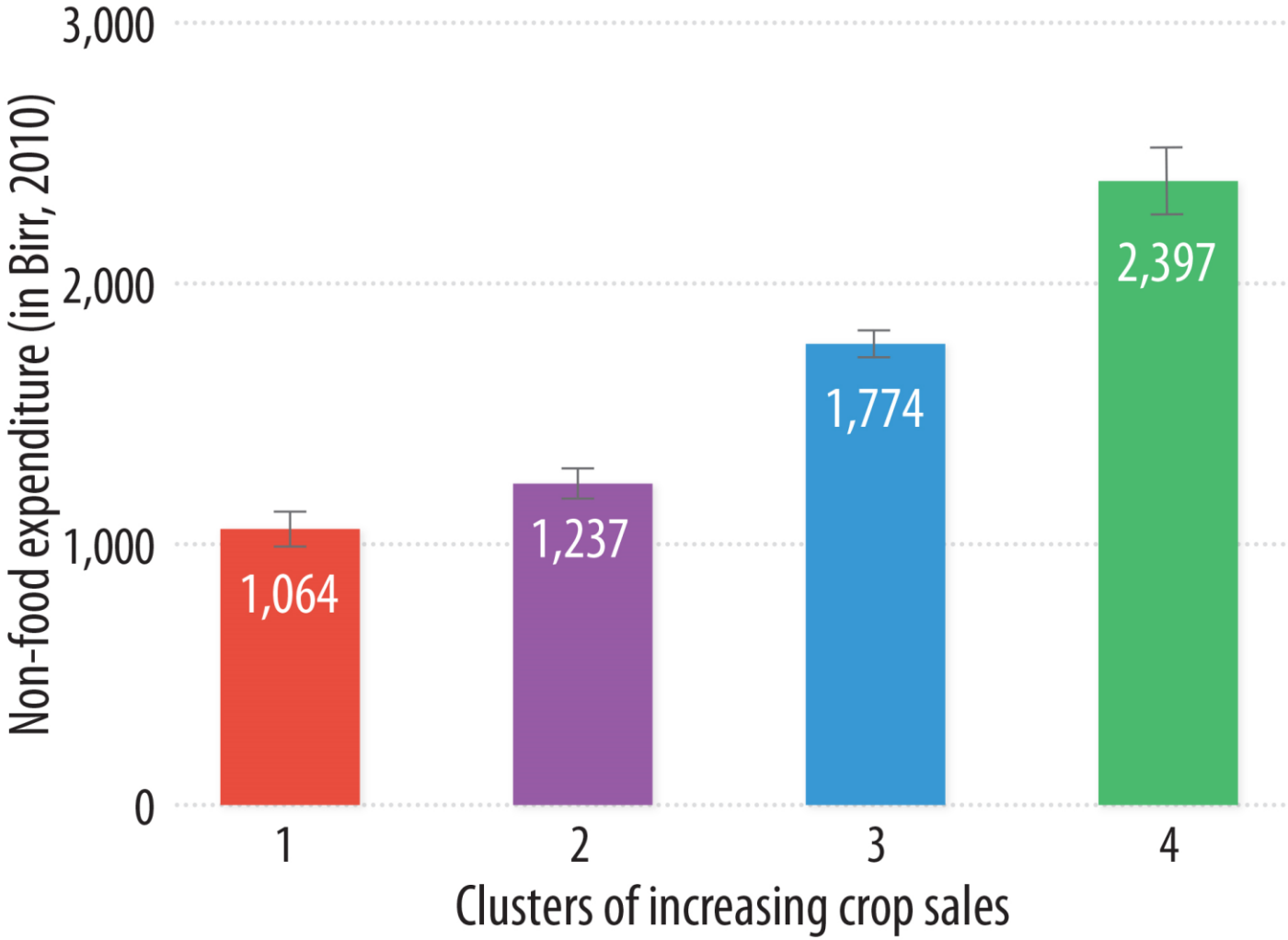
Third, variation in outcomes

Variation in households without food deficiencies across clusters



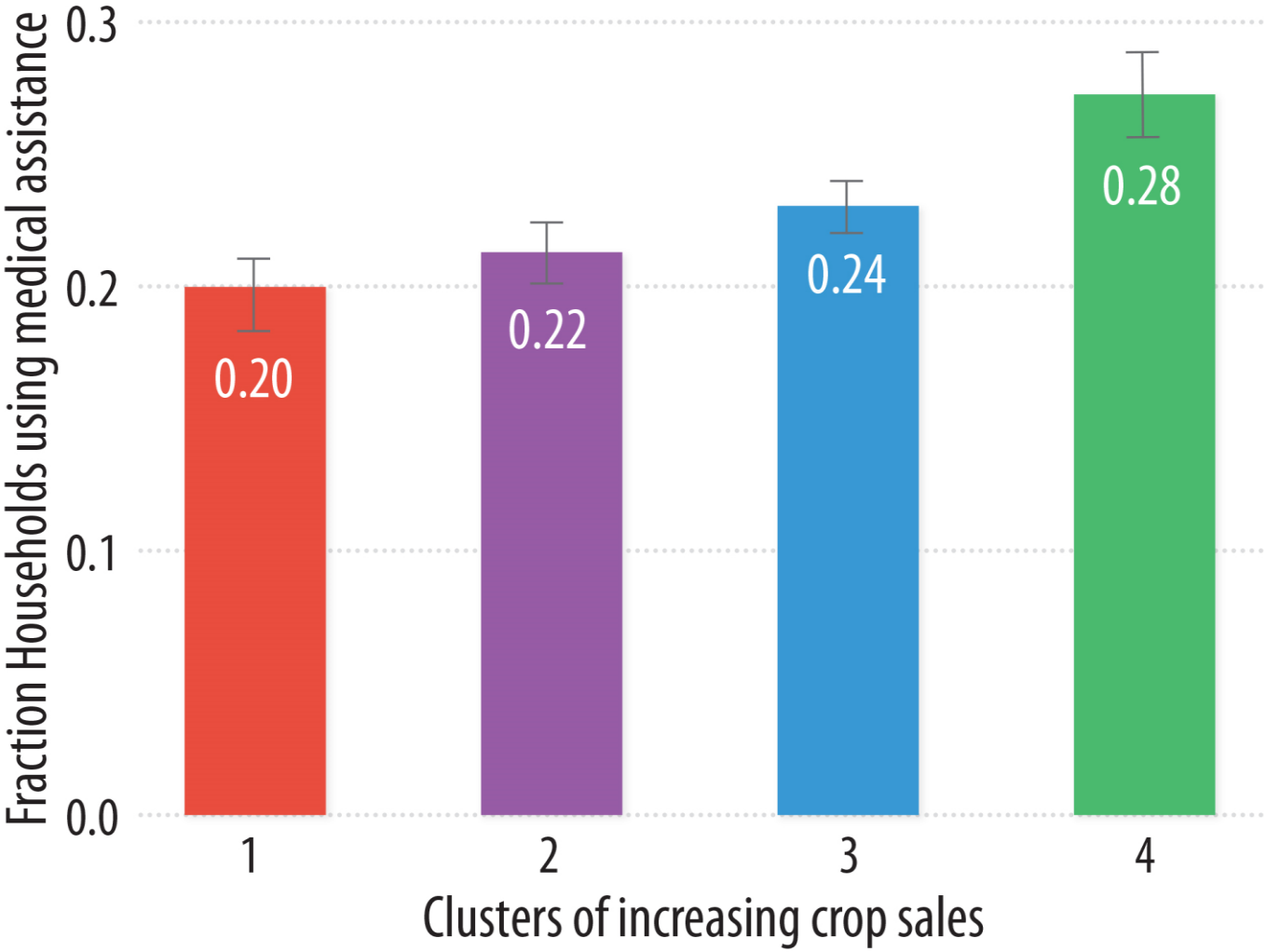
WORLD BANK GROUP

The variation in non-food expenditure across clusters



WORLD BANK GROUP

Variation in % of households using medical assistance across clusters



WORLD BANK GROUP

Initial policy observations

FOR LOW INCOME CLUSTER

Expand equipment (oxen and ploughs) and crop diversification

FOR MIDDLE INCOME CLUSTERS

Improve all the other features

FOR HIGH INCOME CLUSTER

Increase hired workers and increasing savings

Optimizing income in a cluster

WORLD BANK GROUP

How to maximize income within a cluster

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Most Impactful Input….	**Increase farmers' savings**	**Increase # of hired workers**	**Increase # of hired workers**	**Increase # of hired workers**
…which is highly correlated with these inputs	*No other variables*	*No other variables*	*No other variables*	*No other variables*
Input Coefficient	0.134	0.222	0.6	0.292
Input Coefficient Interpretation	Every 32% increase in % of farmers in cluster who saves, associated with 157 Birr increase in average income in cluster	Every 1 additional worker hired is associated with 393 Birr increase in average income in cluster	Every 1 additional worker hired is associated with 1424 Birr increase in average income in cluster	Every 19 additional workers hired is associated with 933 Birr increase in average income in cluster
Other Impactful Input….	**Increase # of oxen owned**	**Obtain Water Storage Pit**	**Increase quantity of chemical fertilizers Used**	**Use extension program**
…. which is highly correlated with these inputs	*Ownership of plough Use of chemical fertiliser*	*No other variables*	*Use of extension program*	*Use of chemical fertilizer*
Input Coefficient	0.106	0.11	0.263	0.054
Input Coefficient Interpretation	Every 0.65 increase in # of oxen owned, associated with 124 Birr increase in average income in cluster	Every 1.1 increase in water storage pit ownership, associated with 195 Birr increase in average income in cluster	Every 1093kg average increase in amount of fertilizer used, associated with 624 Birr increase in average income in cluster	Every 45% increase in usage of extension program within cluster, associated with 172 Birr increase in average income in cluster

Pathway analysis

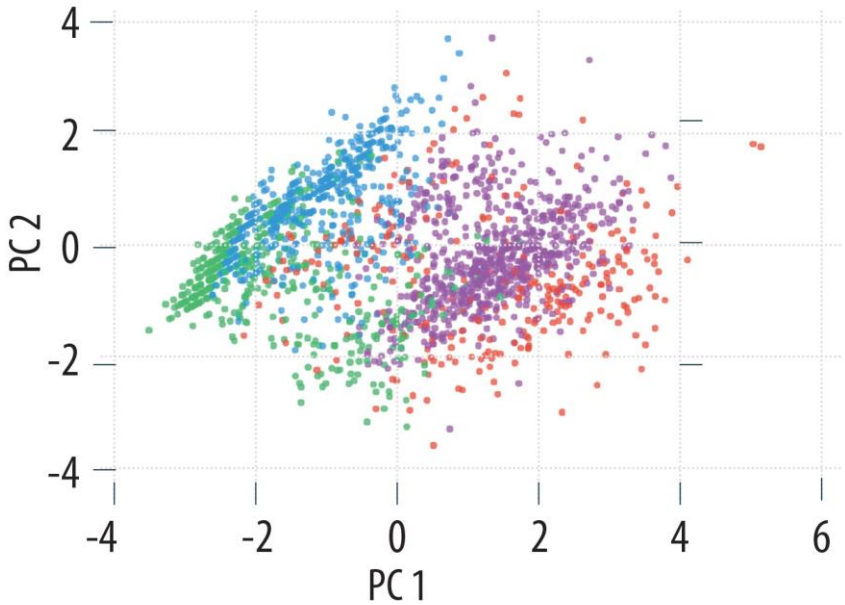
WORLD BANK GROUP

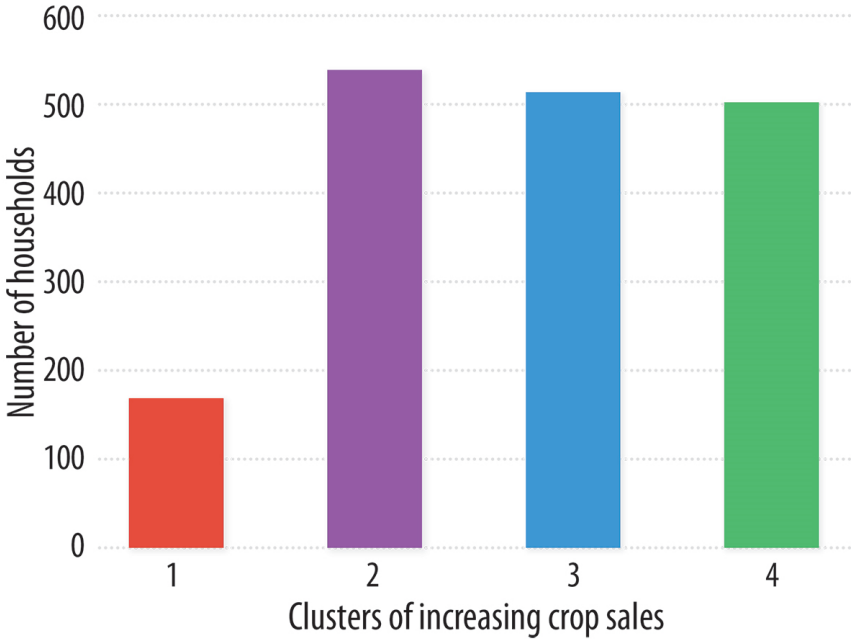
Which pathways do we actually observe?

	CLUSTER 1	CLUSTER 2	CLUSTER 3
Rate of moving over time: % Households that moved to a higher cluster (from 2011 to 2013 or 2013 to 2015)	**23.6%**	**32.9%**	**17.6%**
1st most impactful input (from optimisation analysis)	Has saved	Number of hired workers	Number of hired workers
LIFT FACTOR 1: By how much an increase in input will be associated with an increase in the probability of moving to a higher cluster	No temporal data available (only collected for 2015 wave)	Farmers in this cluster who increase the hired number of workers have a **34%** higher probability of moving to a higher cluster	Farmers in this cluster who increase the hired number of workers have a **32%** higher probability of moving to a higher cluster
Other impactful input (also from optimisation analysis)	Number of oxen owned	Number of water storage pit owned	Quantity of chemical fertilizers used
LIFT FACTOR 2: By how much an increase in input will be associated with an increase in the probability of moving to a higher cluster	Farmers in this cluster who increase the hired number of workers have a **7%** higher probability of moving to a higher cluster	Farmers in this cluster who acquire more water storage pits have a **18%** higher probability of moving to a higher cluster	Farmers in this cluster who increase the chemical fertilizers that they use have a **12%** higher probability of moving to a higher cluster

TANZANIA: Results overview

K-means clustering results

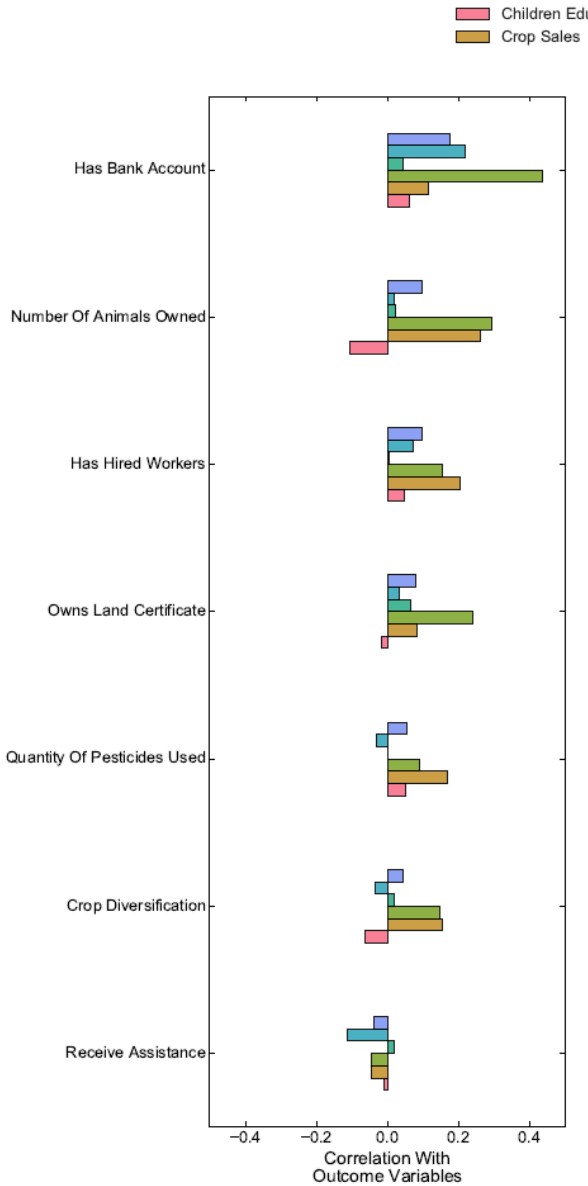




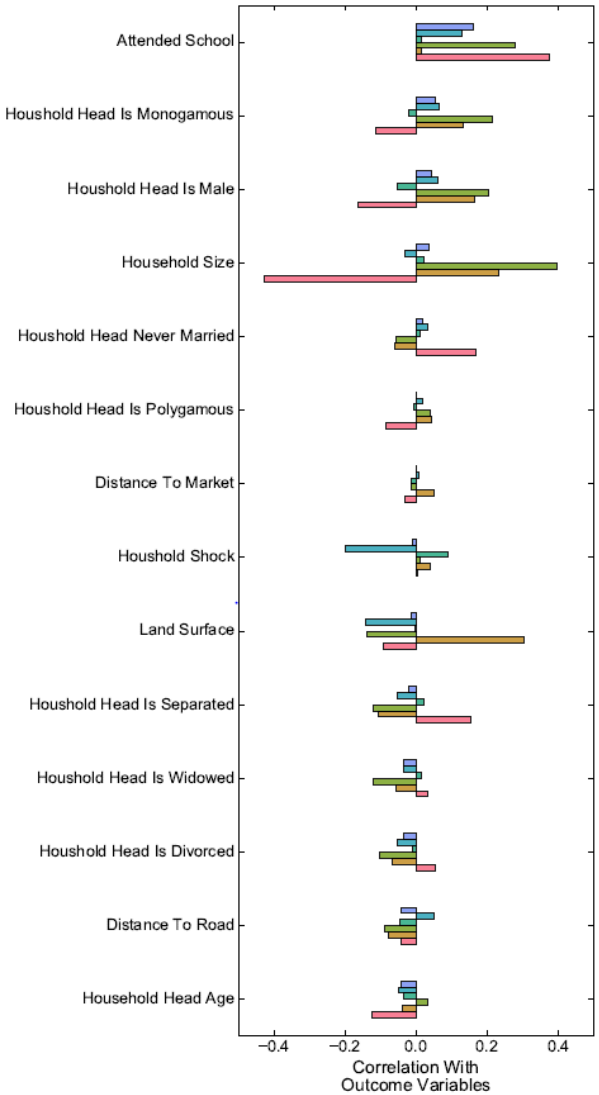
- K-means clustering achieves desired outcome: clusters farmers based on their own unique set of actionable variables most correlated with outcomes and not with other input variable
- Clustering consistent over time
- We pick: number of clusters = **4**

Correlation between Tanzania inputs and outcomes

- Correlations between inputs and outcomes are of lower order of magnitude as Ethiopia's correlations
- Strength of correlations for individual inputs are not the same for the 2 countries: point to possible different pathways and optimized input variables



Actionable inputs



Non-actionable inputs

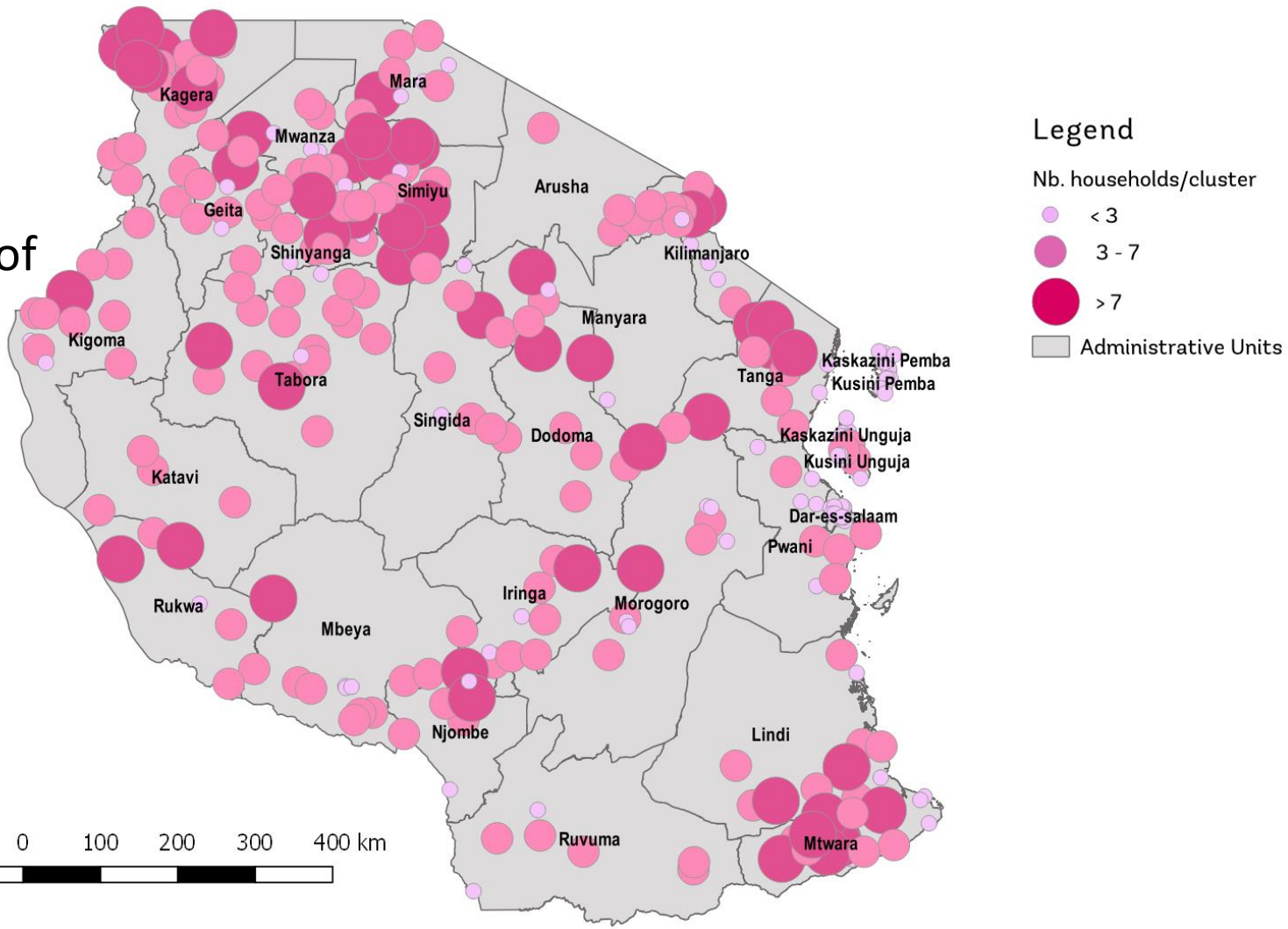
WORLD BANK GROUP

List of most correlated inputs for k-means clustering of households in Tanzania's LSMS-ISA data

- Whether the farmer has hired workers (but not the number of workers hired)
- The number of animals owned
- Quantity of pesticides used
- Crop diversification
- Household size
- Whether the household head is separated

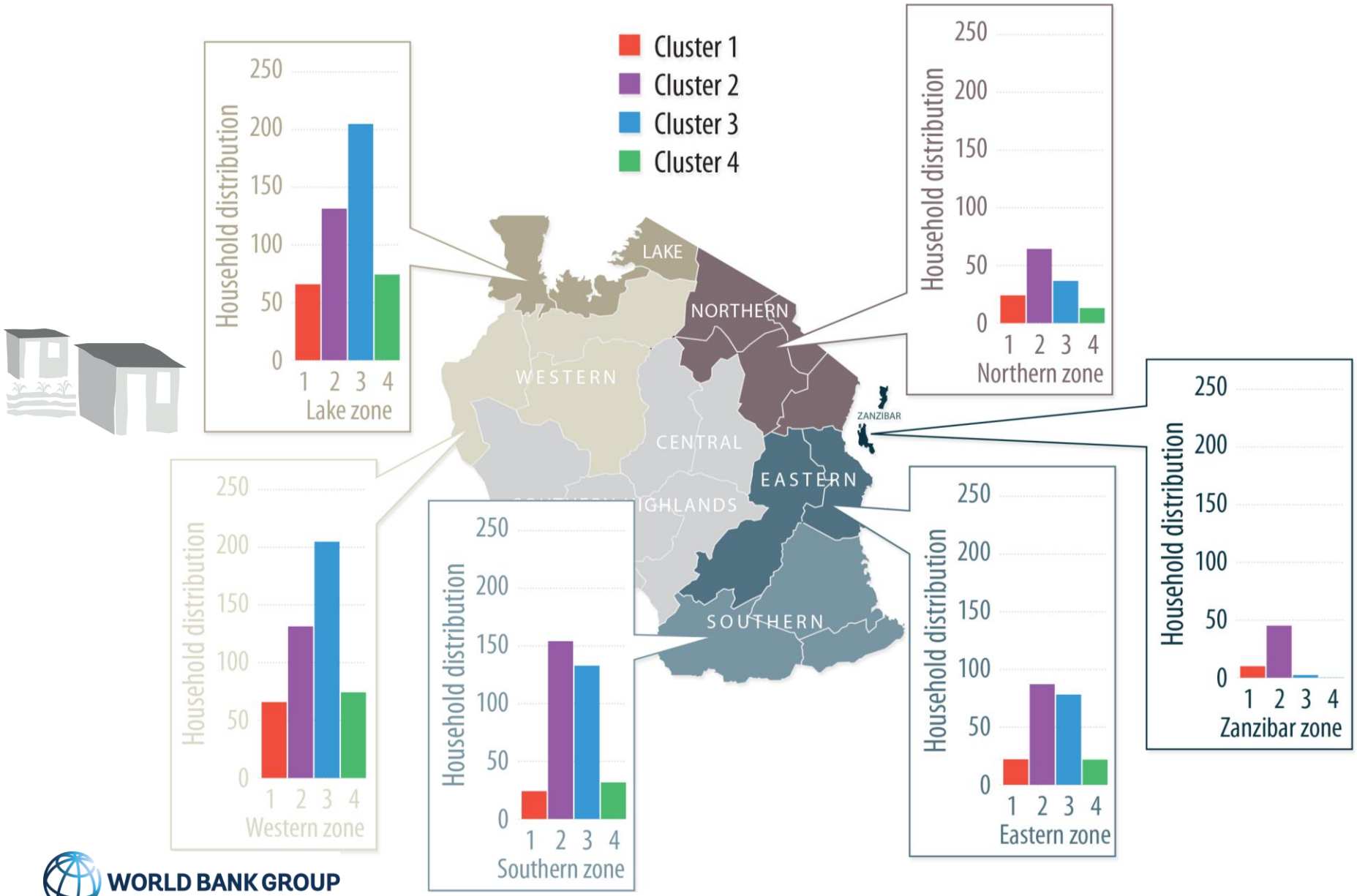
Where are households in the clusters?

3 to 8 households in close geographic proximity have similar GPS coordinates to protect confidentiality of panel survey participants, hence the heat map approach



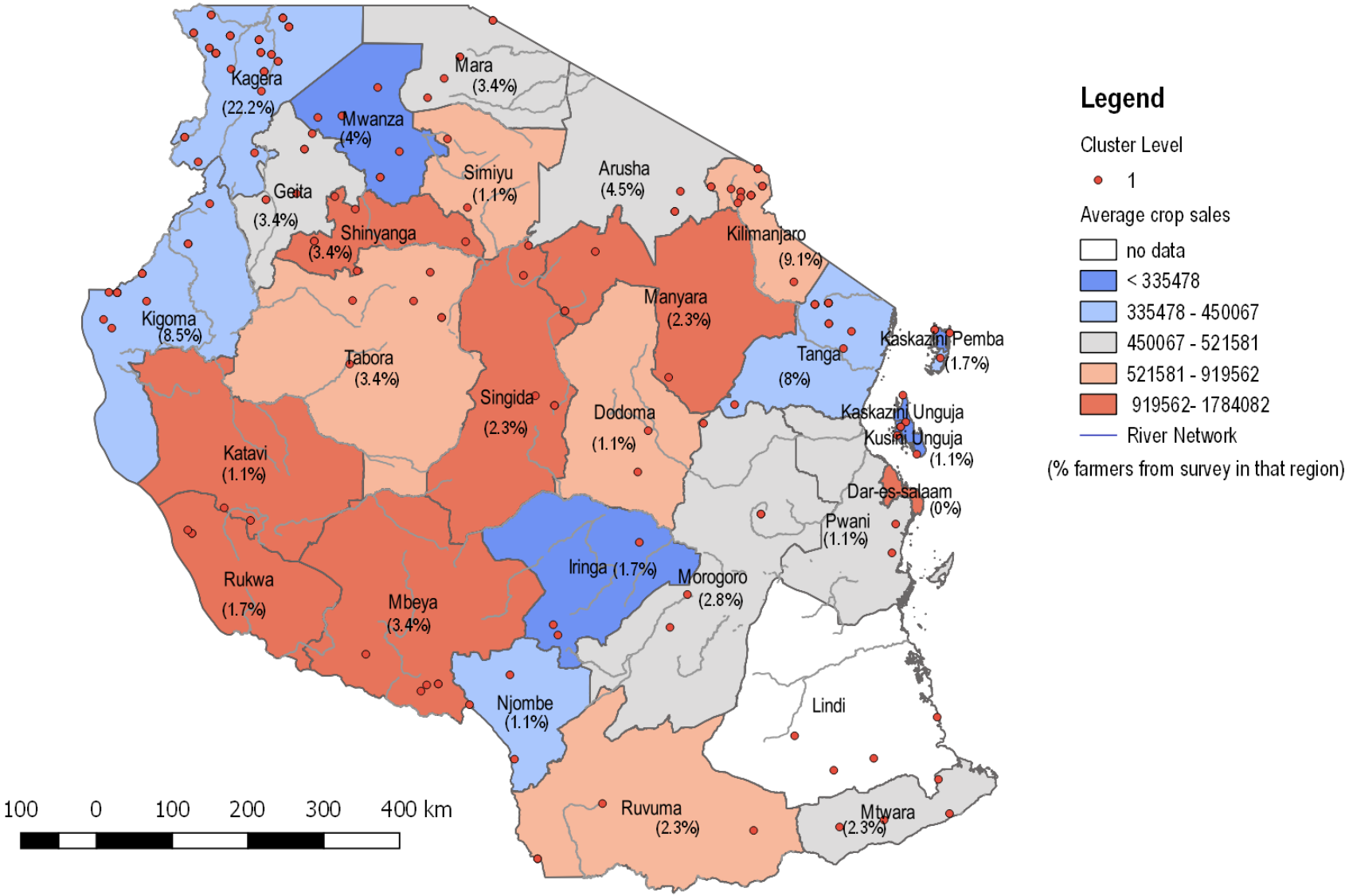
WORLD BANK GROUP

Household Distribution by Cluster and Administrative Zone



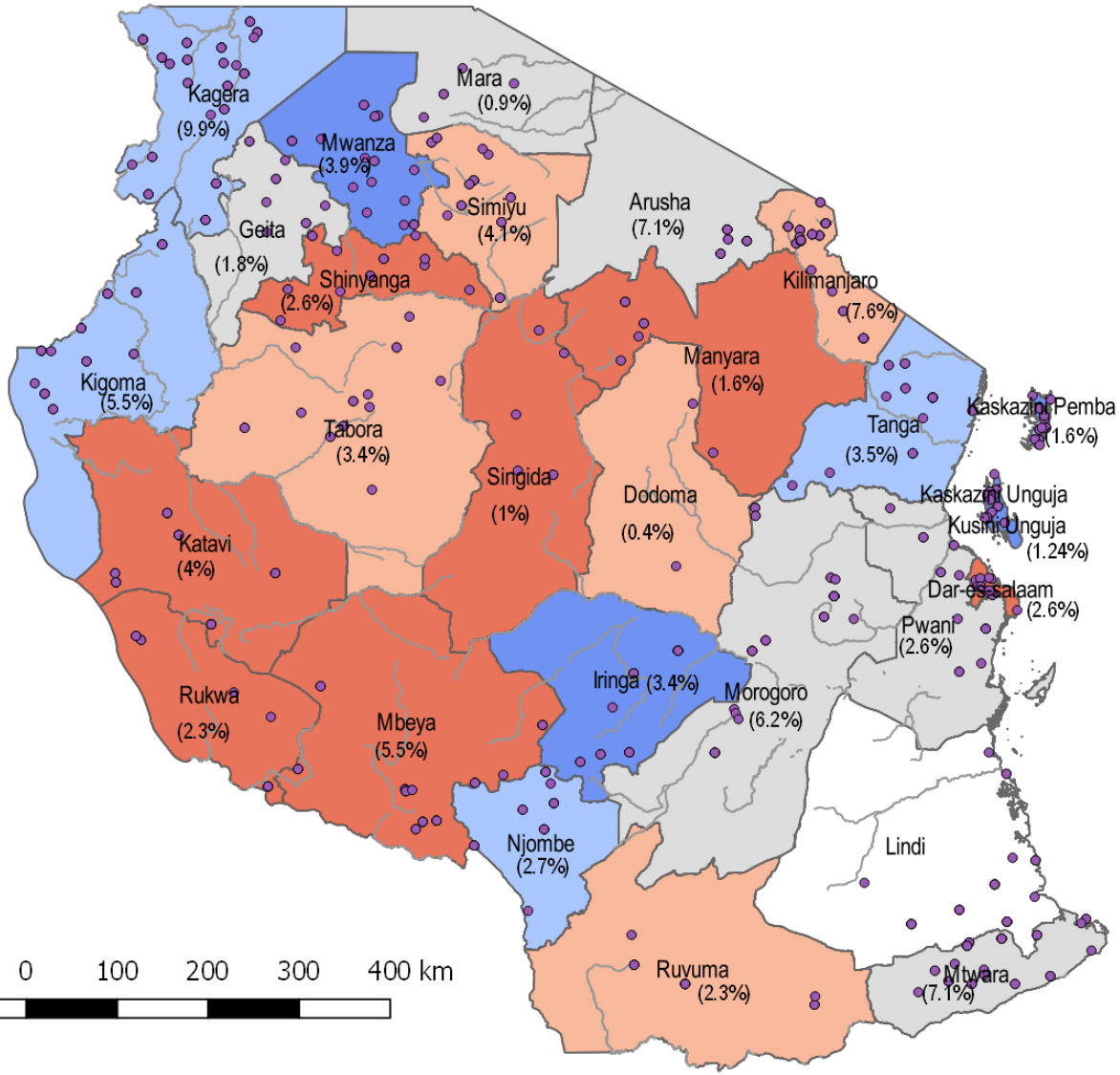
WORLD BANK GROUP

Household Distribution by Cluster and Administrative Zone



WORLD BANK GROUP

Household Distribution by Cluster and Administrative Zone



Legend

Cluster Level

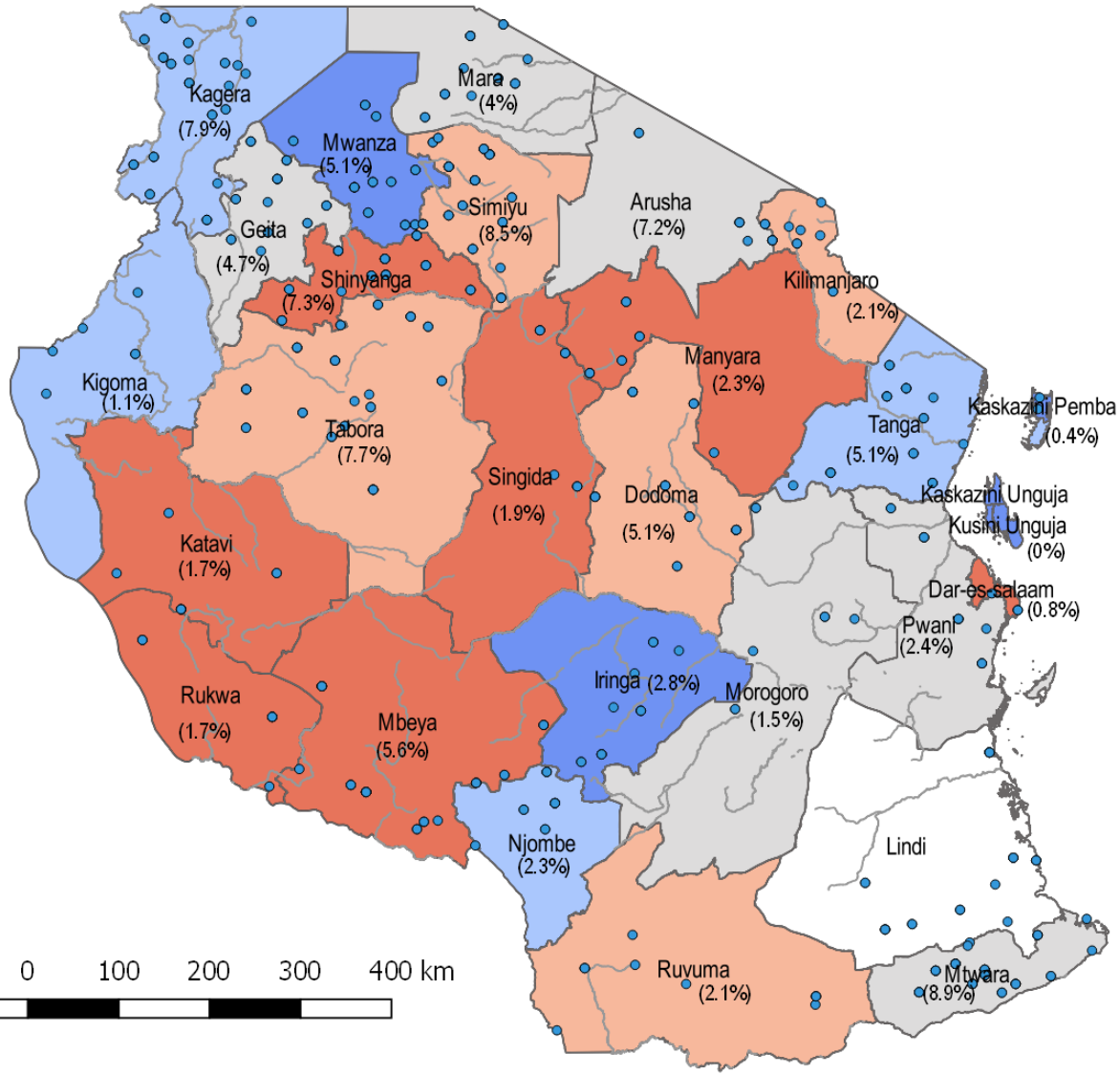
- 2

Average crop sales

- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562- 1784082
- River Network

WORLD BANK GROUP

Household Distribution by Cluster and Administrative Zone



Legend

Cluster Level

- 3

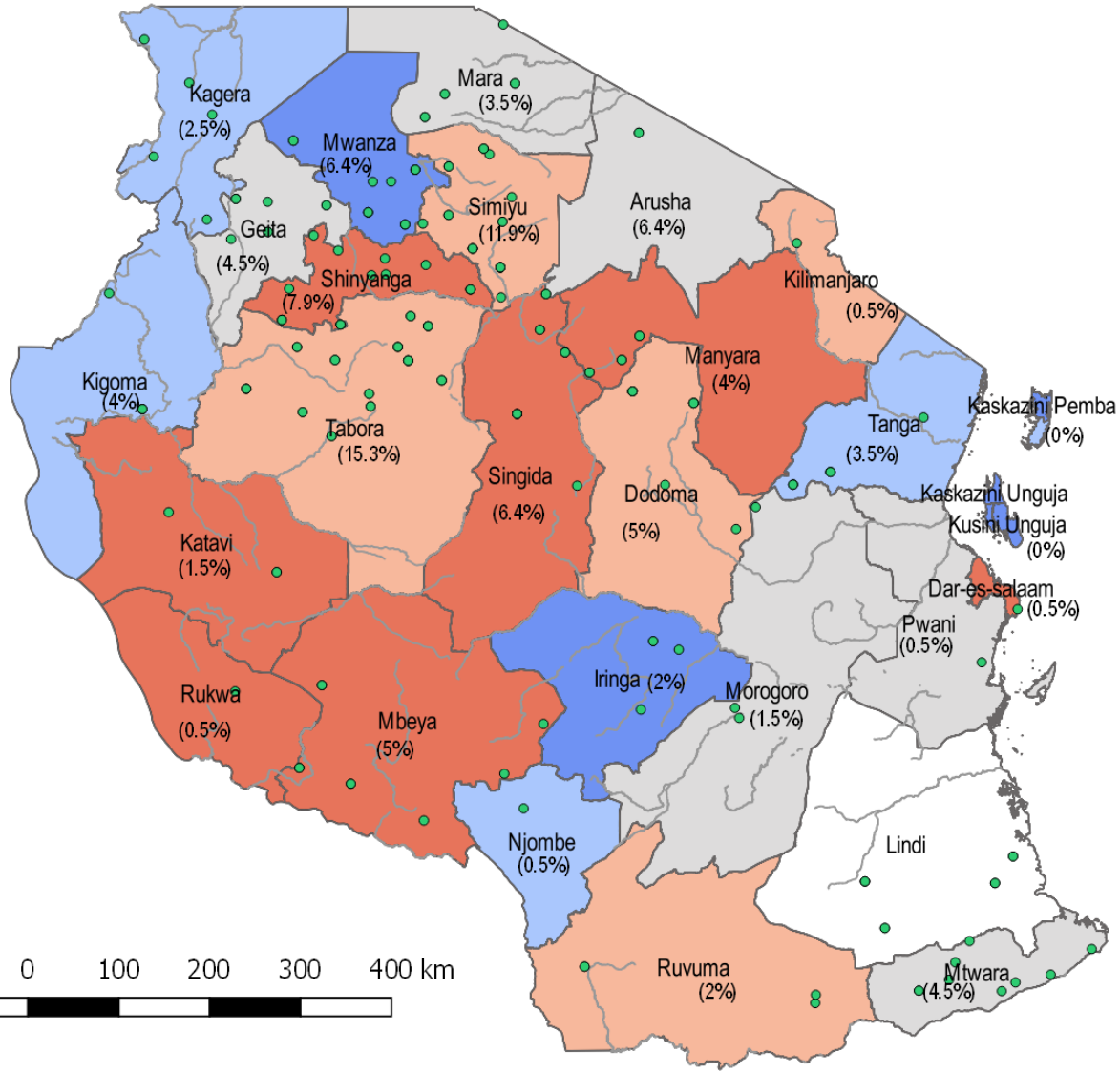
Average crop sales

- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562- 1784082
- River Network

(% farmers from survey in that region)

WORLD BANK GROUP

Household Distribution by Cluster and Administrative Zone



Legend

Cluster Level

- 4

Average crop sales

- no data
- < 335478
- 335478 - 450067
- 450067 - 521581
- 521581 - 919562
- 919562- 1784082
- River Network

(% farmers from survey in that region)

WORLD BANK GROUP

How are the clusters different? *(actionable variables)*

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Number of Crops Planted	6.792	4.929	5.377	4.282
% of Households with Bank Account	0.087	0.116	0.083	0.127
% of Households Who Have Borrowed	0.357	0.34	0.502	0.467
Has Hired Workers	0.563	0.794	0.839	0.883
Number of Months Hiring Workers	0.213	0.543	0.562	0.837
Number Of Animals Owned	5.688	45.117	45.755	75.885
Number Of Hired Workers	10.674	11.378	10.628	15.272
Number Of Ploughs Owned	0.064	0.094	0.212	0.713
% of Households with Land Certificate	0.034	0.076	0.07	0.117
Quantity Of Fertilizers Used (in Kg)	179	281	519	983
Quantity Of Pesticides Used (in Liters)	1820	2520	2532	3698
% of Households Using Credit Services	0.138	0.185	0.195	0.212
% of Households Using Irrigation	0.033	0.052	0.021	0.023

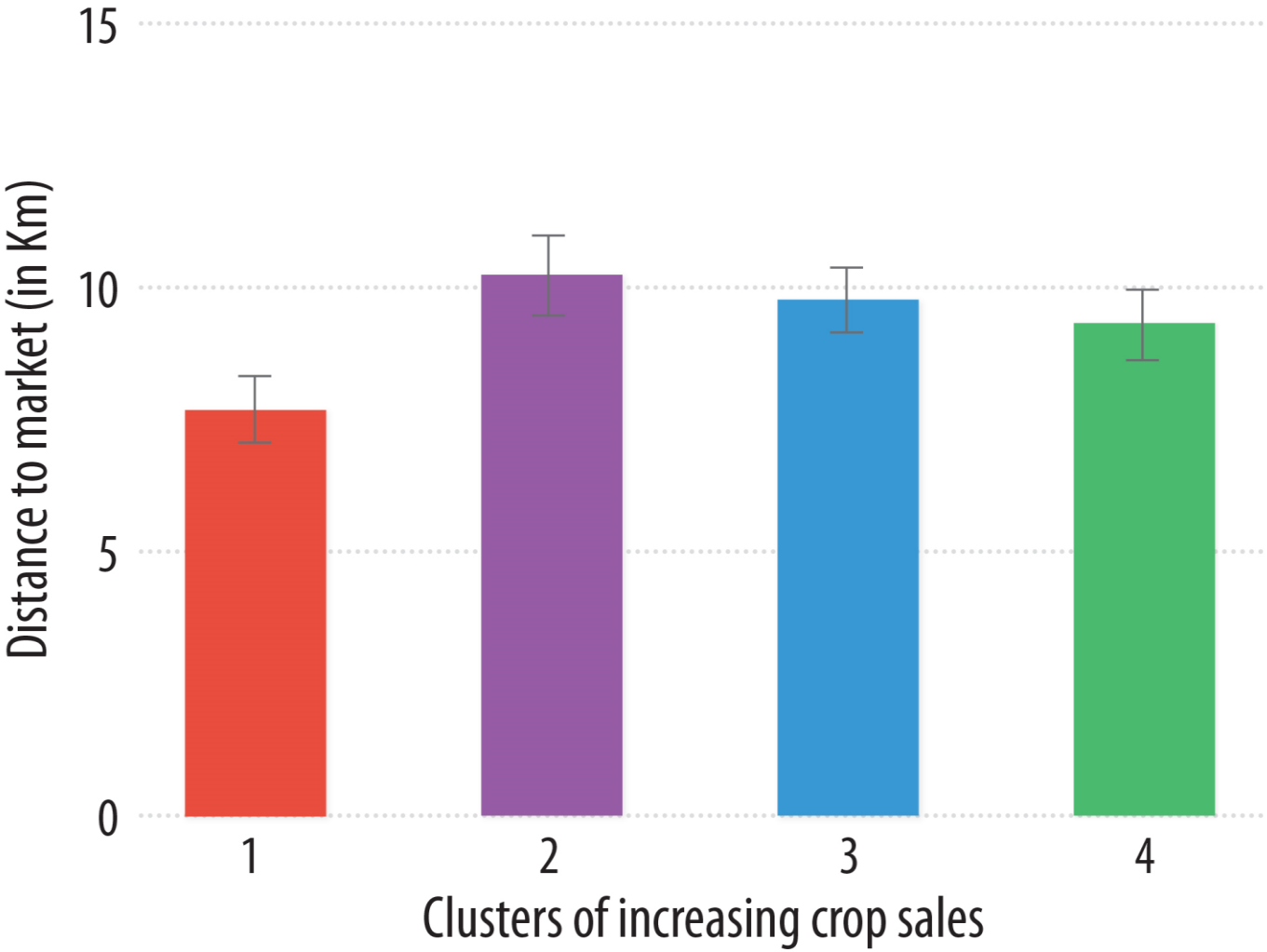
How are the clusters different? (outcomes)

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Children Education	0.406	0.407	0.411	0.398
Crop Sales (in Shilling)	268,759.4	520,175.9	683,827.8	151,234.9
Expenditure (in Shilling)	295,253.3	319,095.6	347,855.5	536,351.6
% of Households Using Medical Assistance	0.294	0.274	0.253	0.206
% of Households Without Food Deficiencies	0.51	0.53	0.566	0.669

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

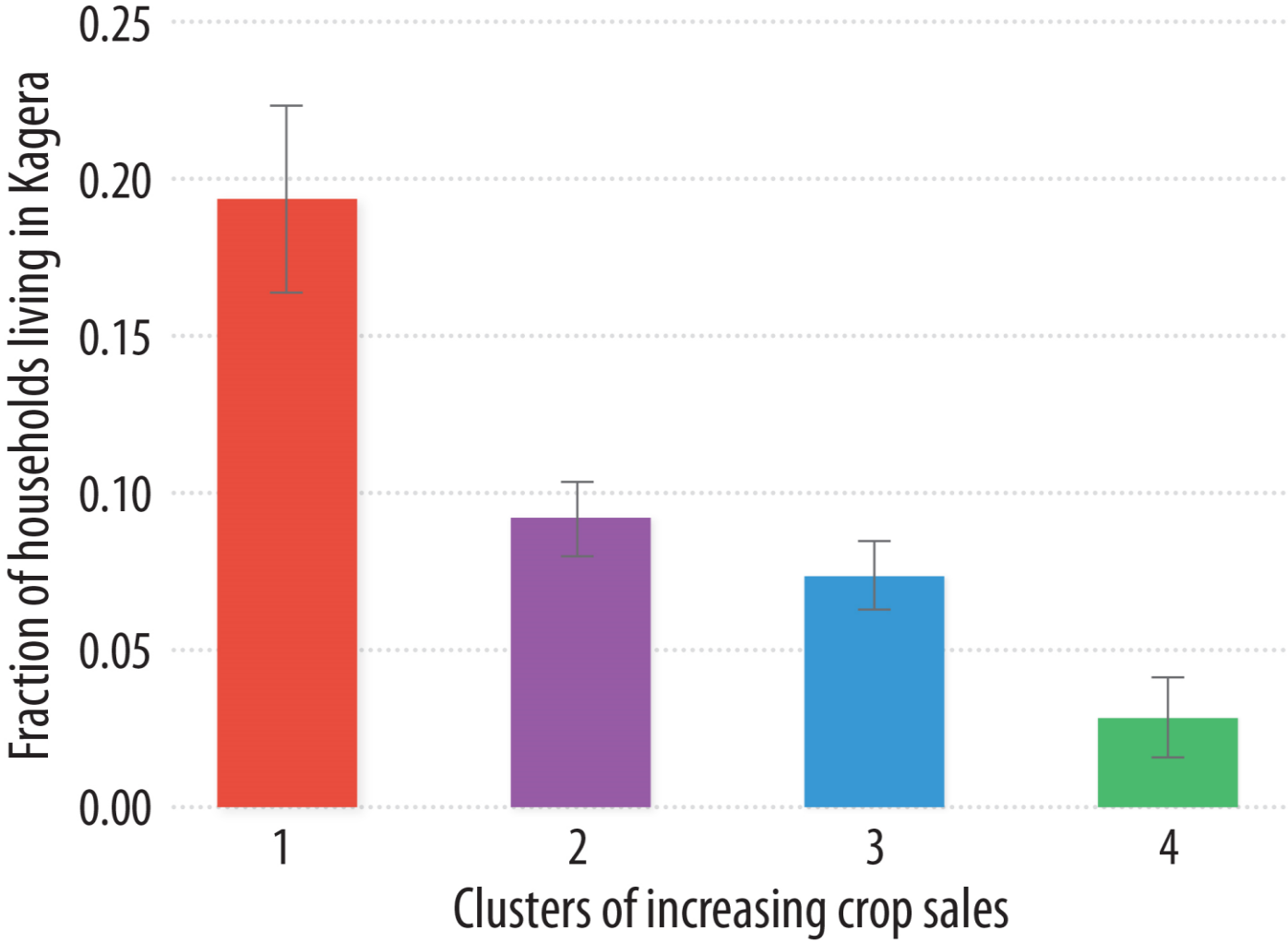
First, variation in non-actionable variables

Variation in distance to market across clusters



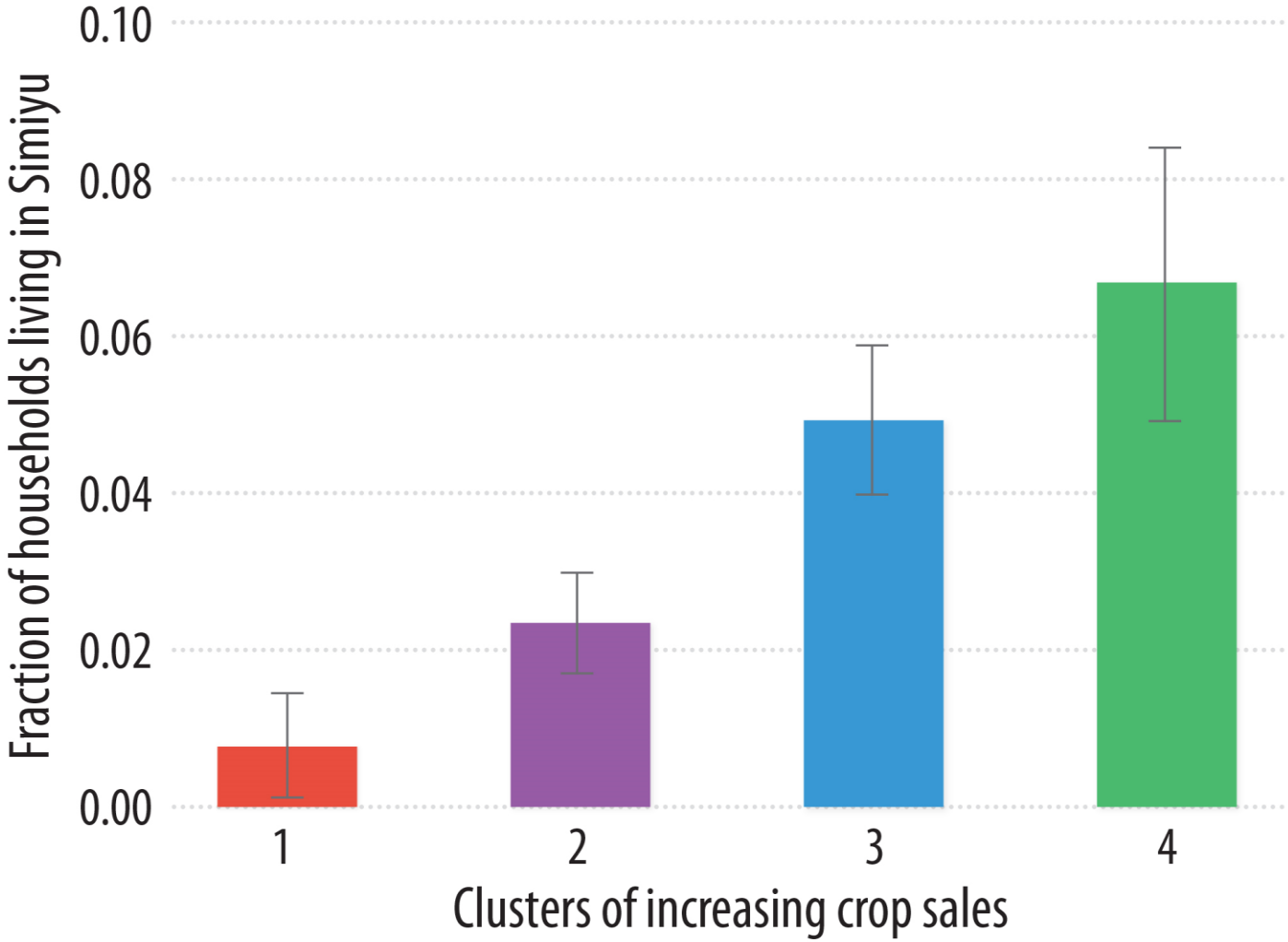
WORLD BANK GROUP

The variation in households living in Kagera across clusters



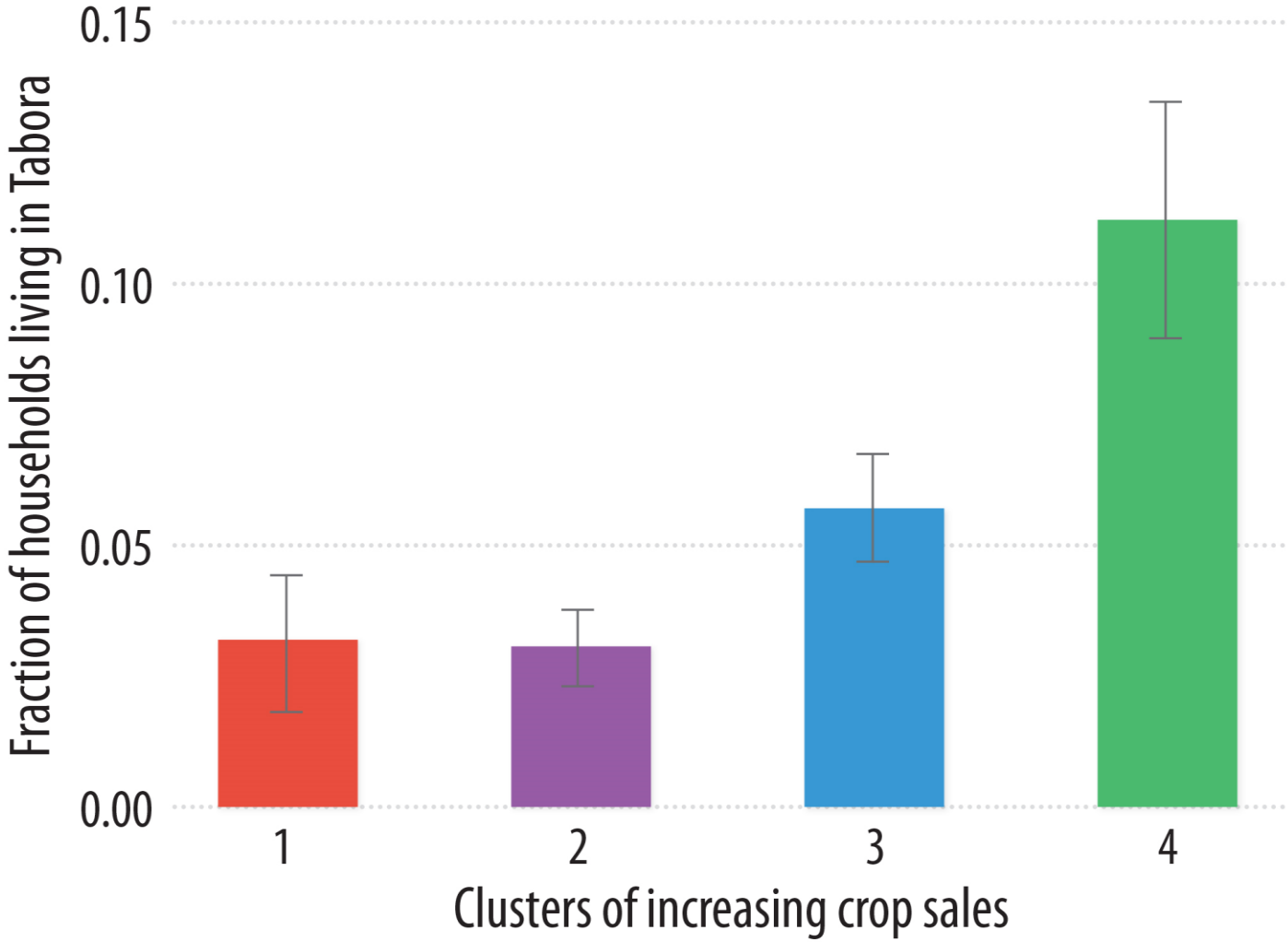
WORLD BANK GROUP

The variation in households living in Simiyu across clusters



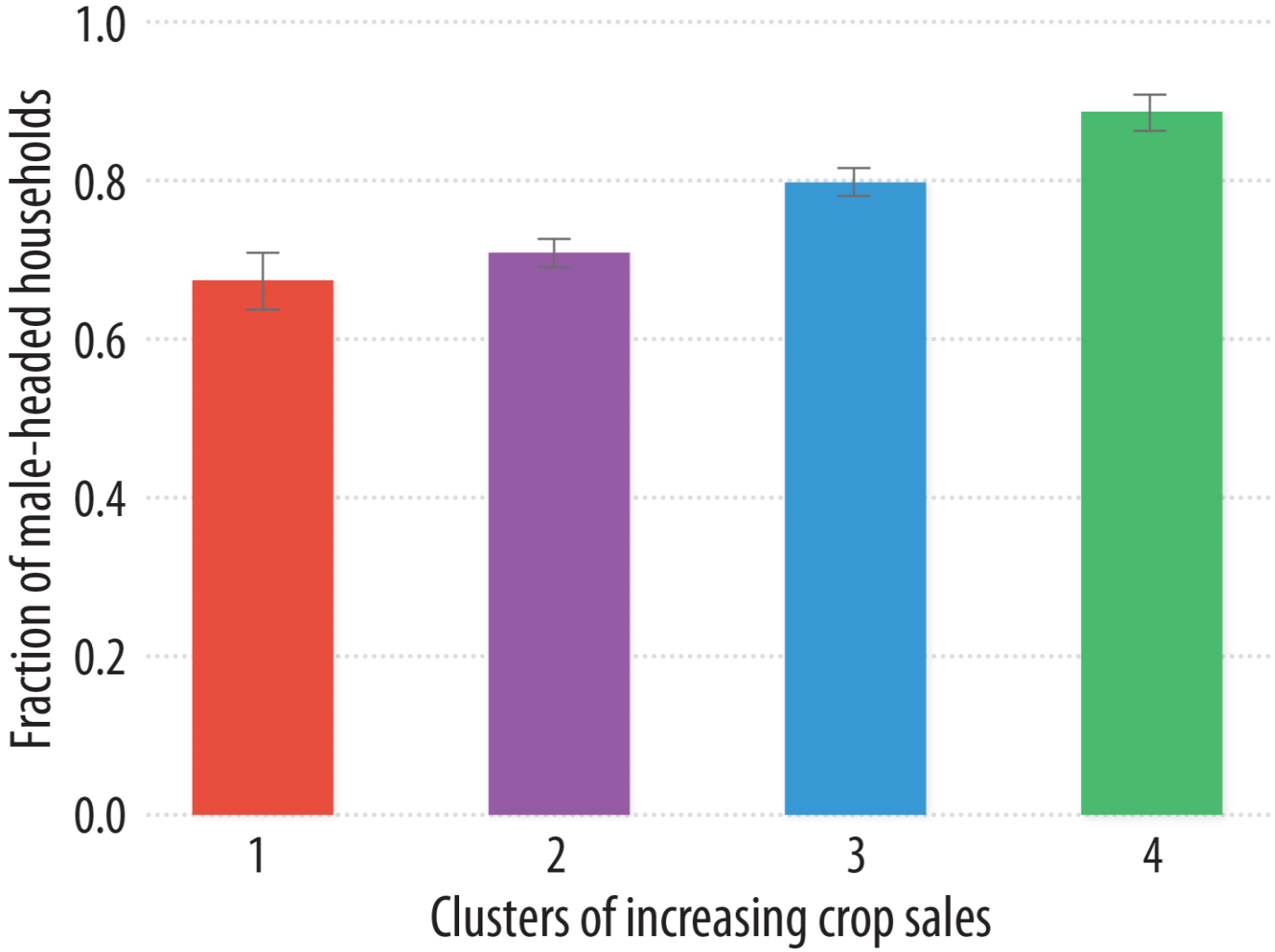
WORLD BANK GROUP

The variation in households living in Tabora across clusters



WORLD BANK GROUP

The variation in % of male-headed households across clusters

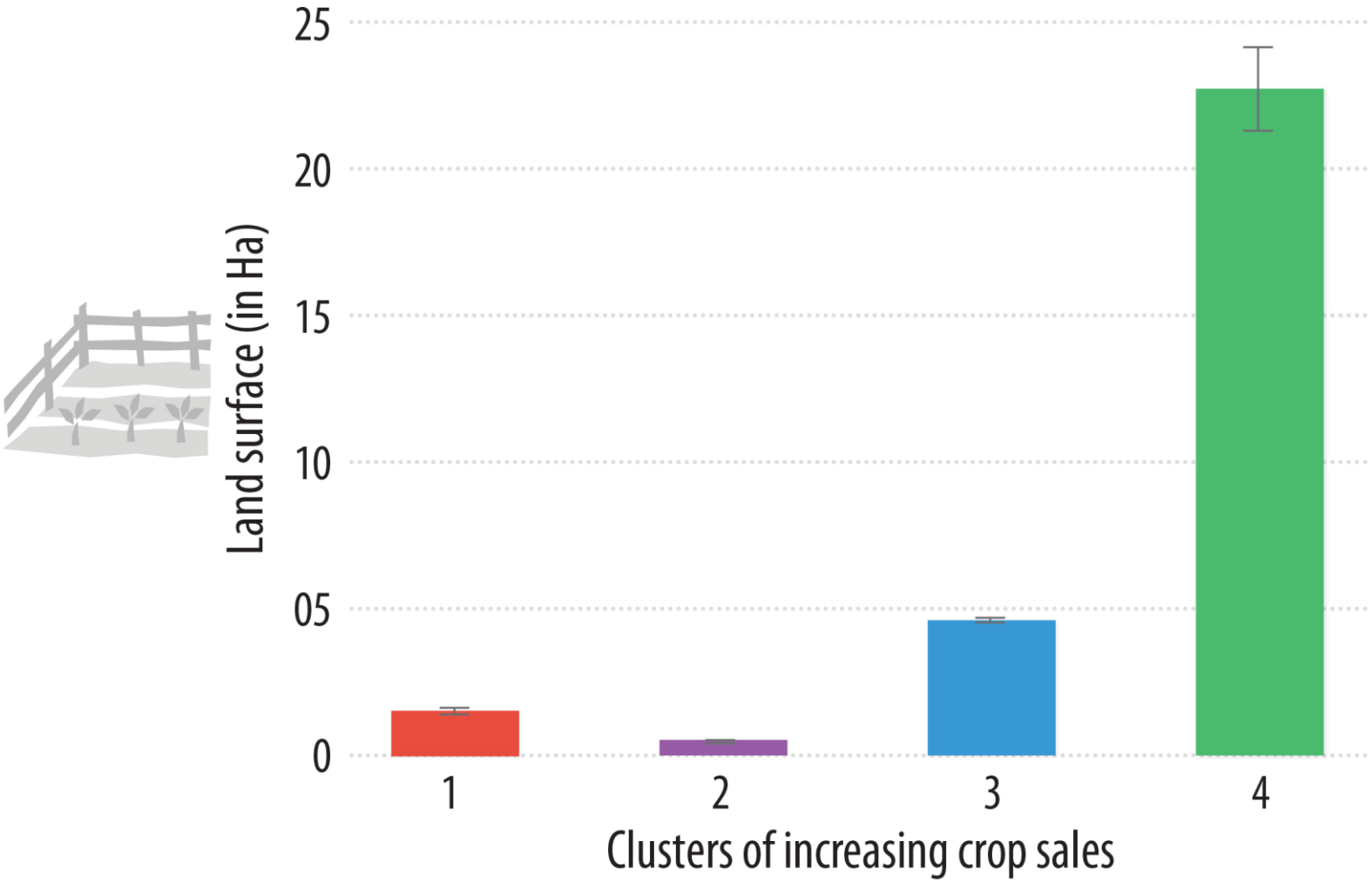


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

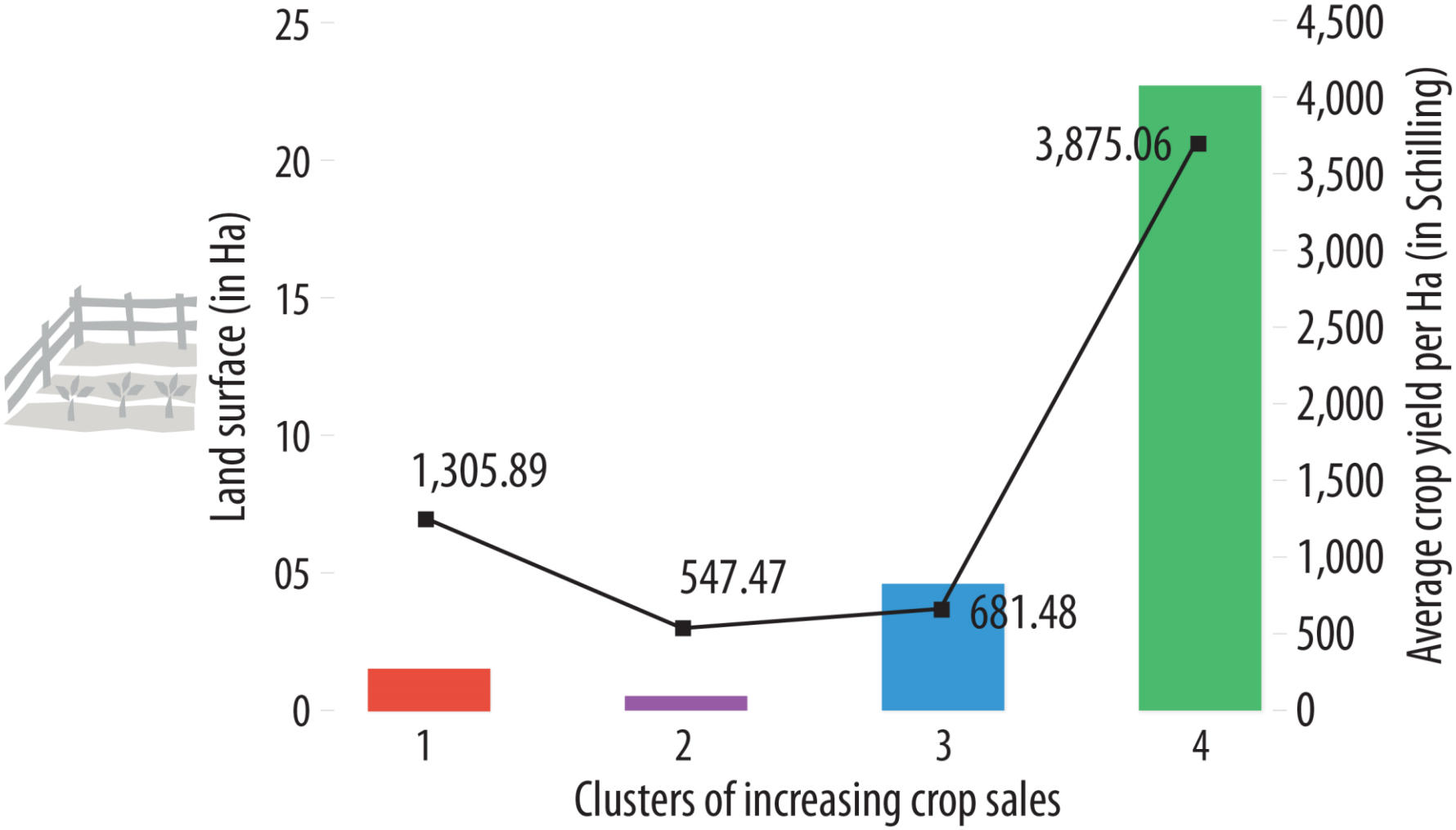
Second, variation in actionable variables

The variation in land surface across clusters



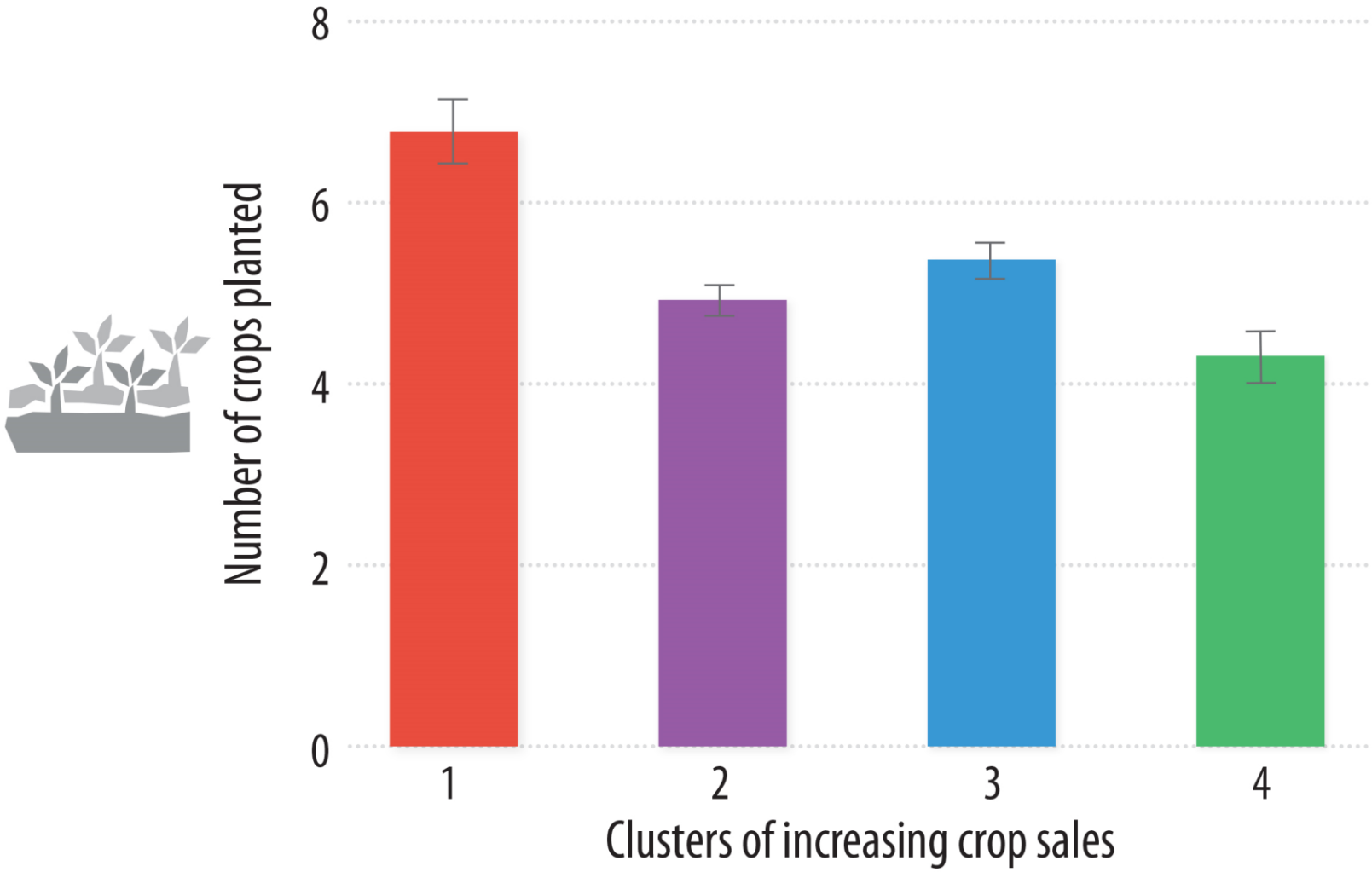
WORLD BANK GROUP

The variation in land surface across clusters



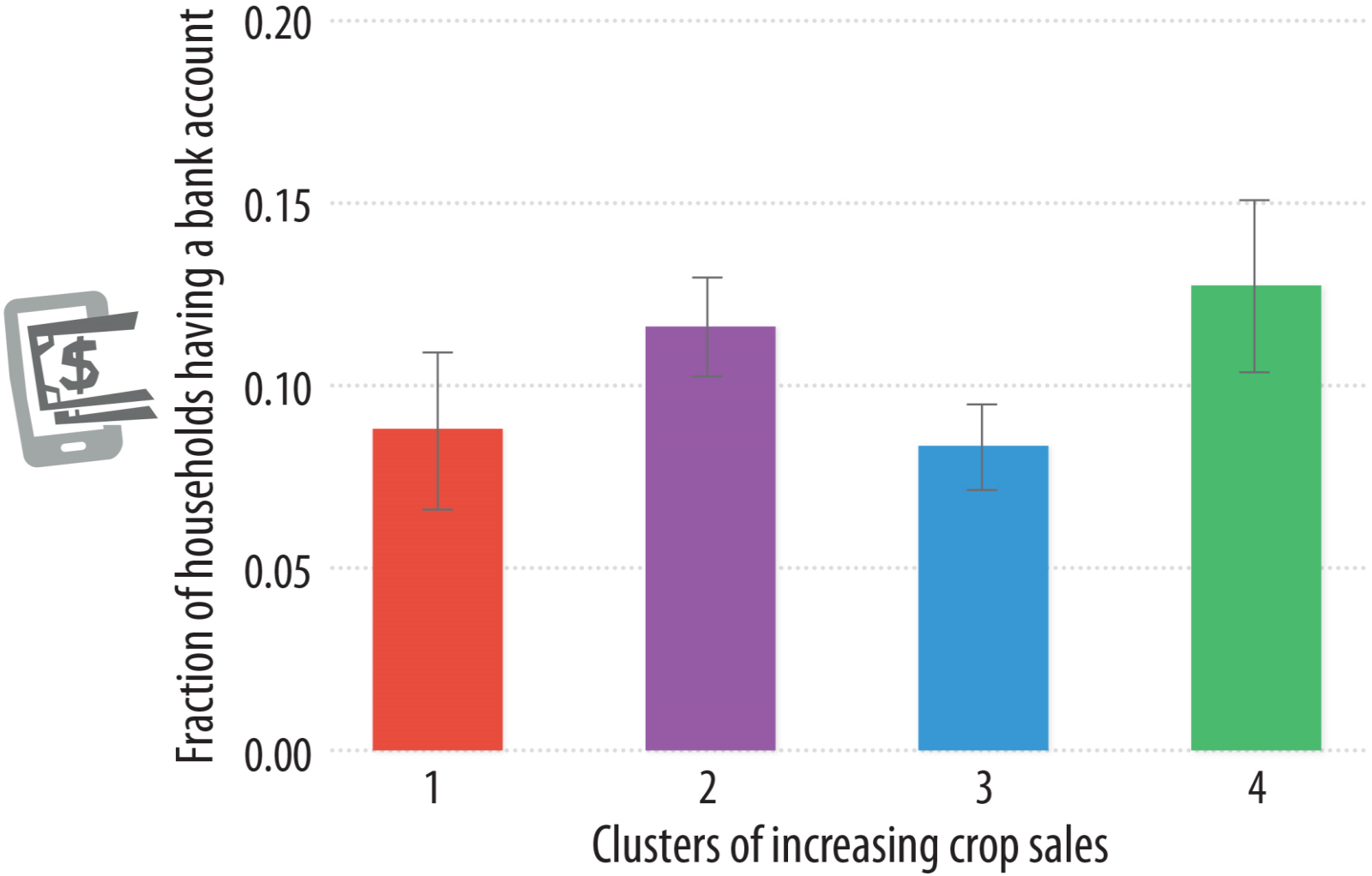
WORLD BANK GROUP

The variation in number of crops planted across clusters



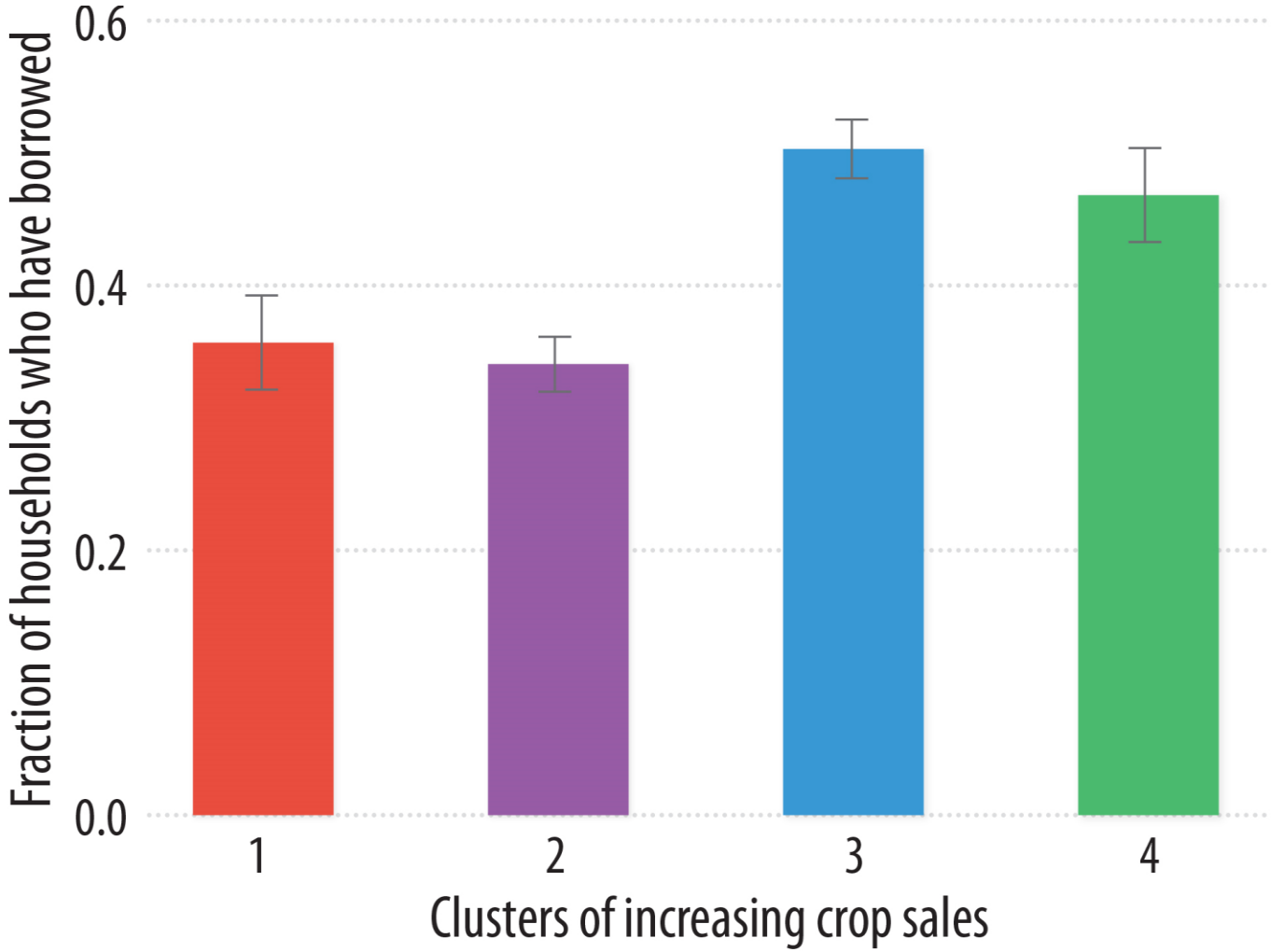
WORLD BANK GROUP

The variation in % of households having a bank account across clusters



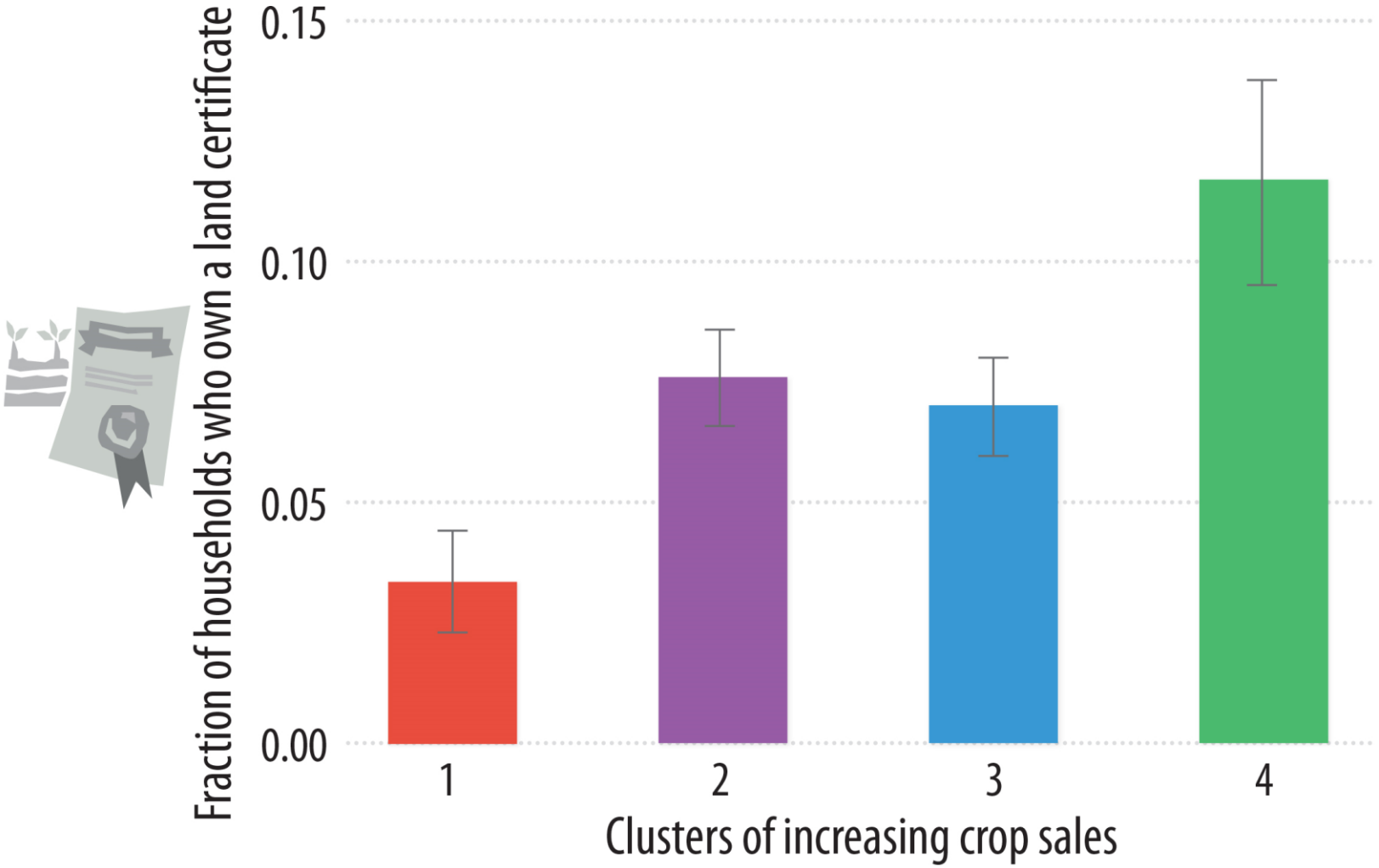
WORLD BANK GROUP

The variation in % of households who have borrowed across clusters



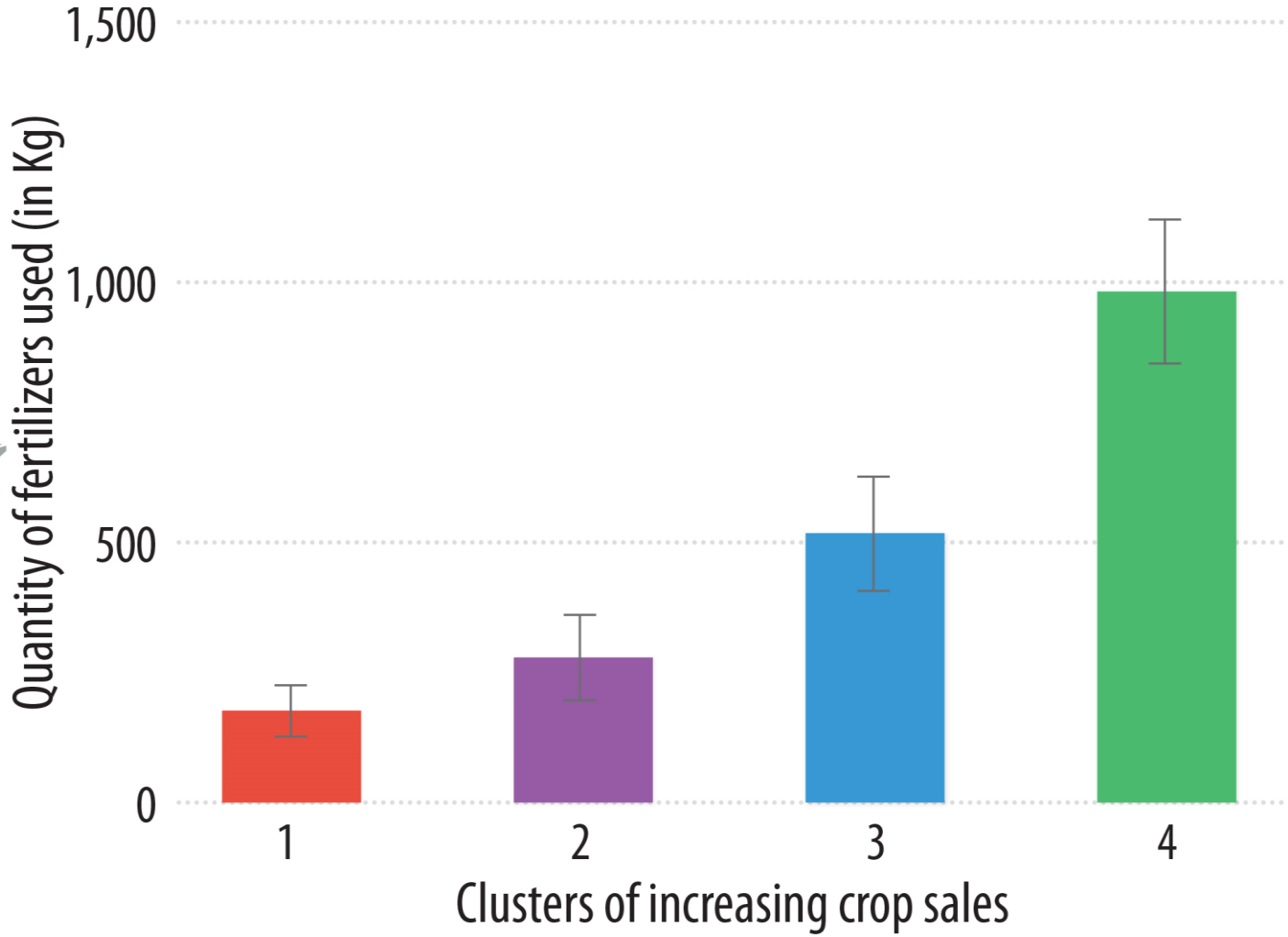
WORLD BANK GROUP

The variation in % of households who own a land certificate across clusters



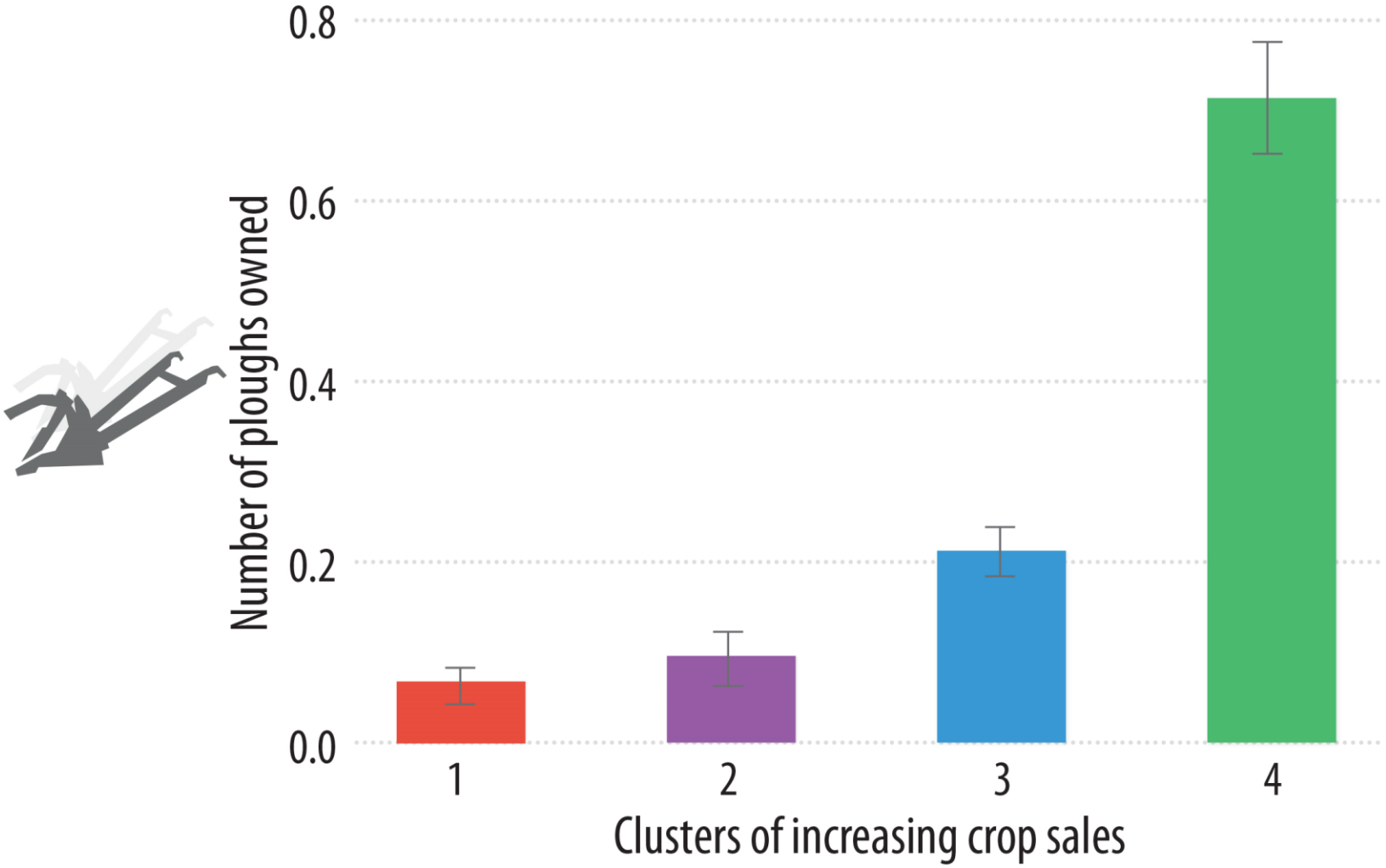
WORLD BANK GROUP

The variation in quantity of fertilizers used across clusters



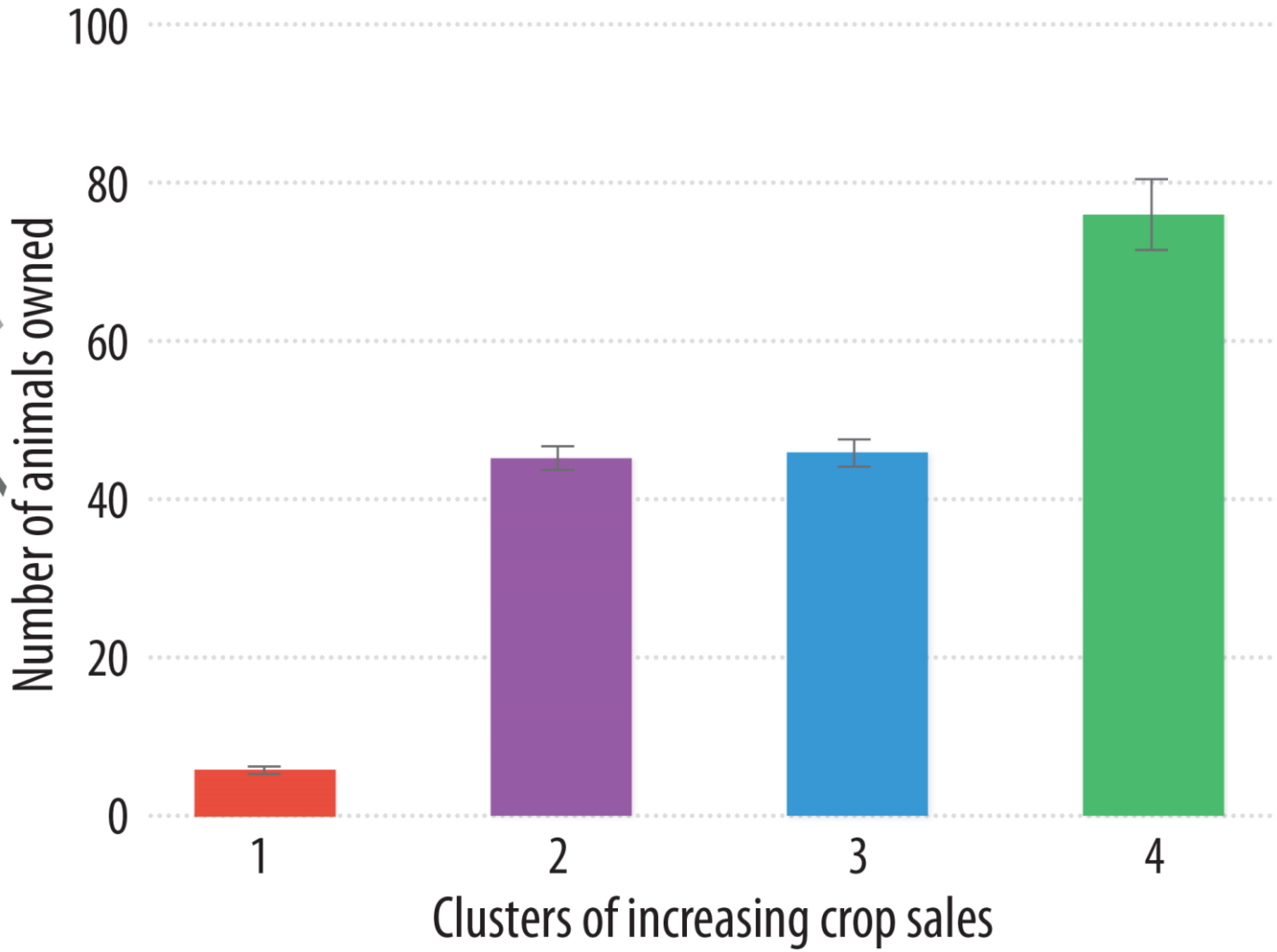
WORLD BANK GROUP

The variation in number of ploughs owned across clusters



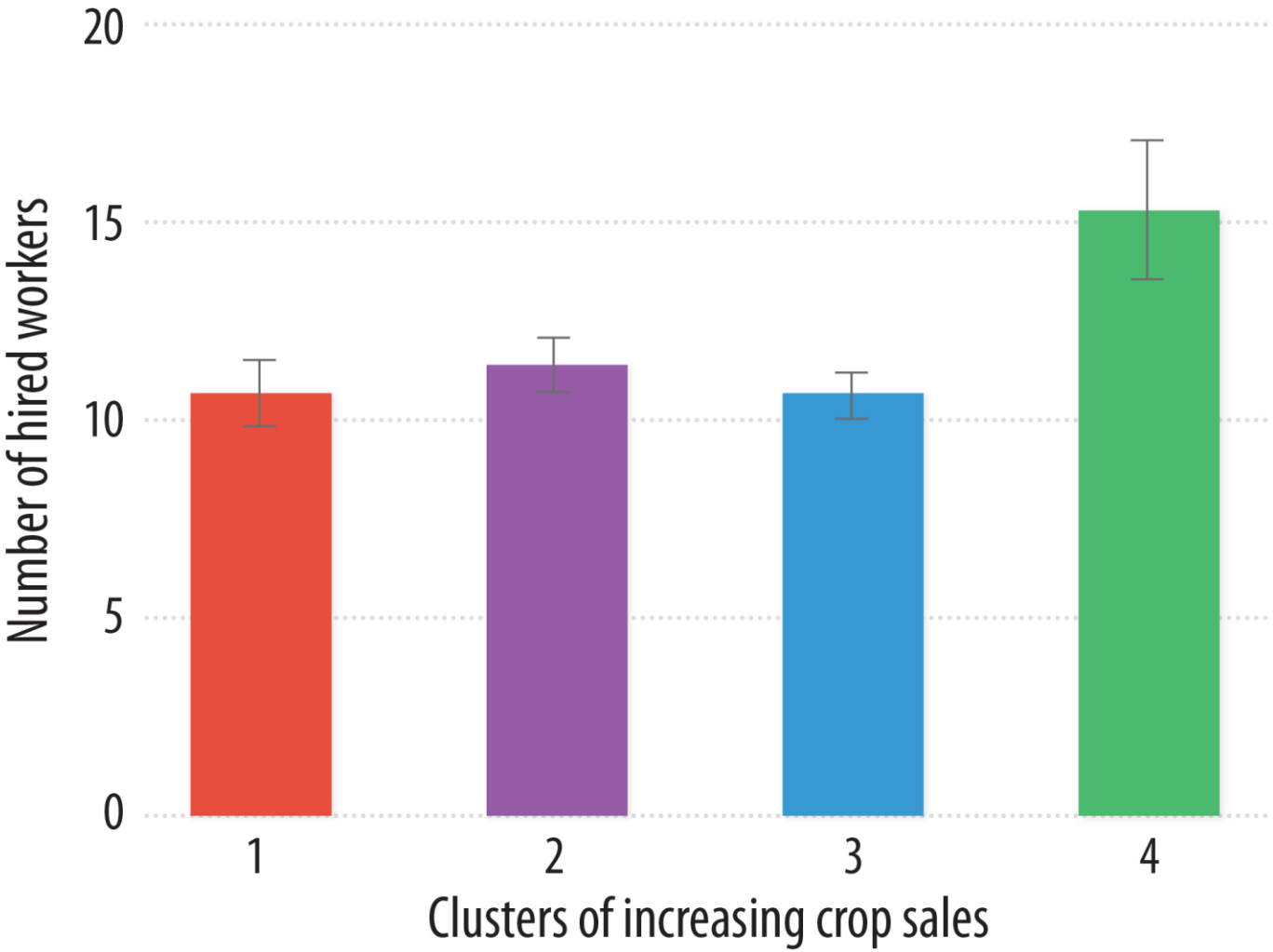
WORLD BANK GROUP

The variation in number of animals owned across clusters



WORLD BANK GROUP

The variation in number of hired workers across clusters

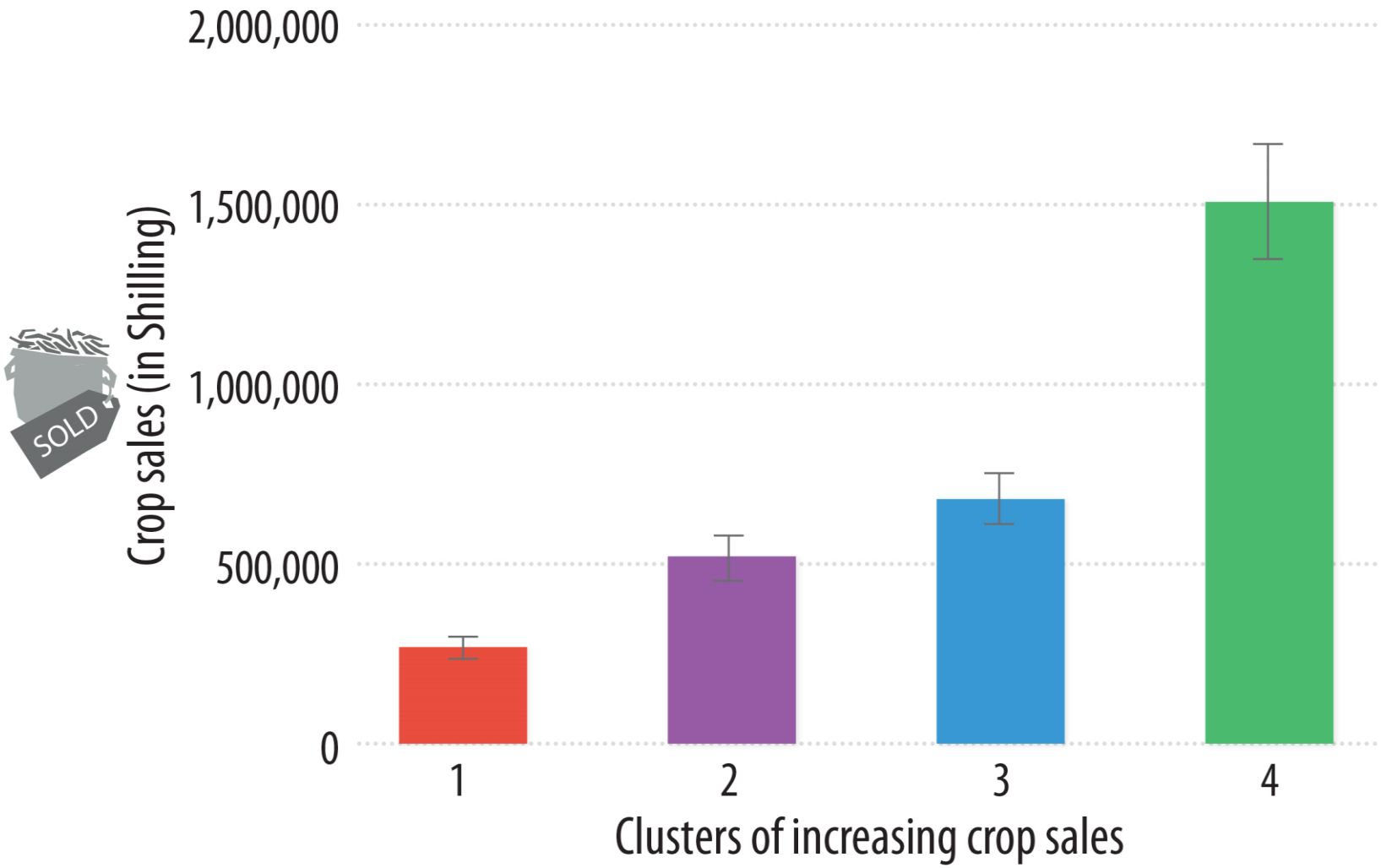


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

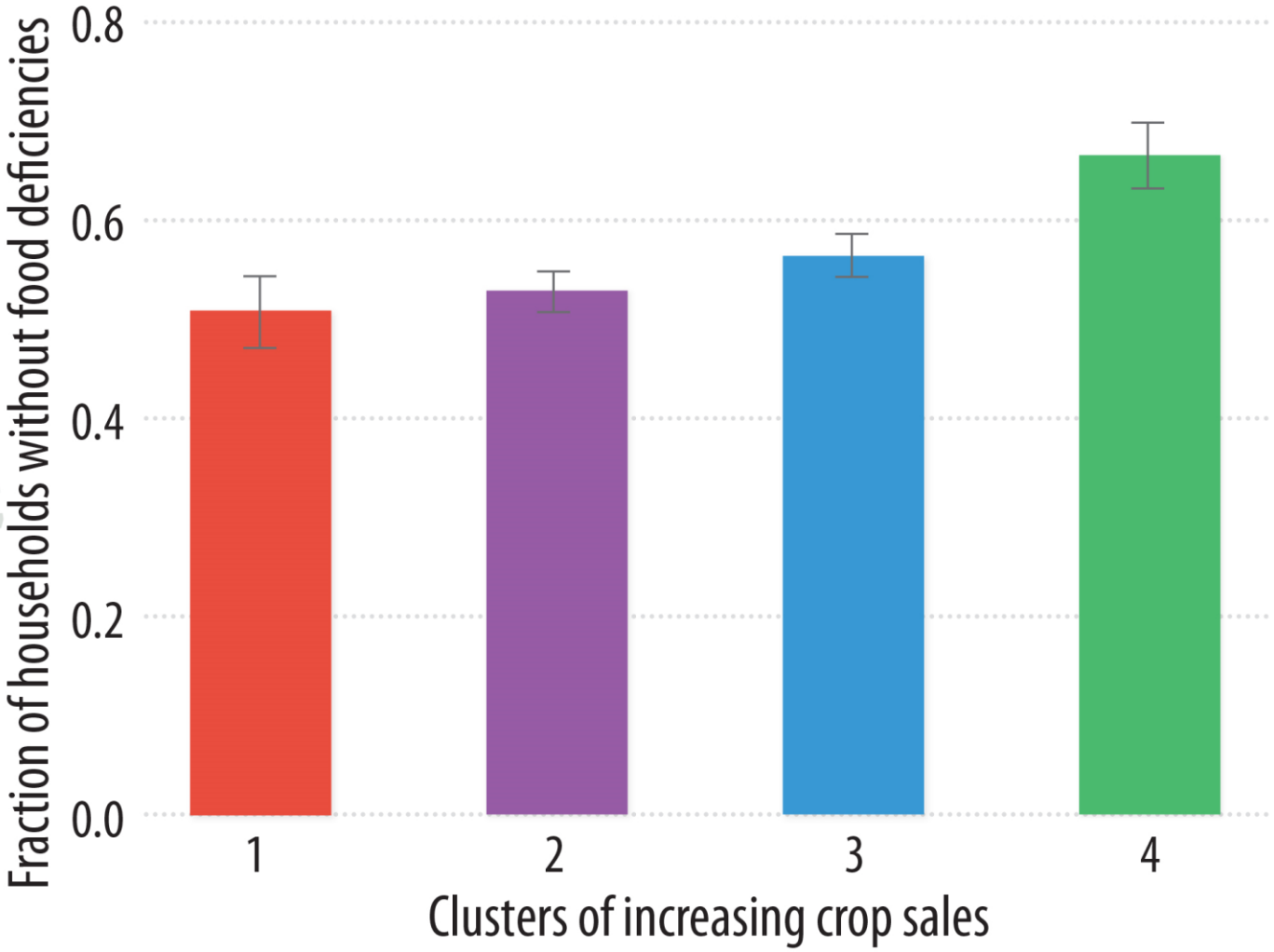
Third, variation in outcomes

The variation in crop sales across clusters



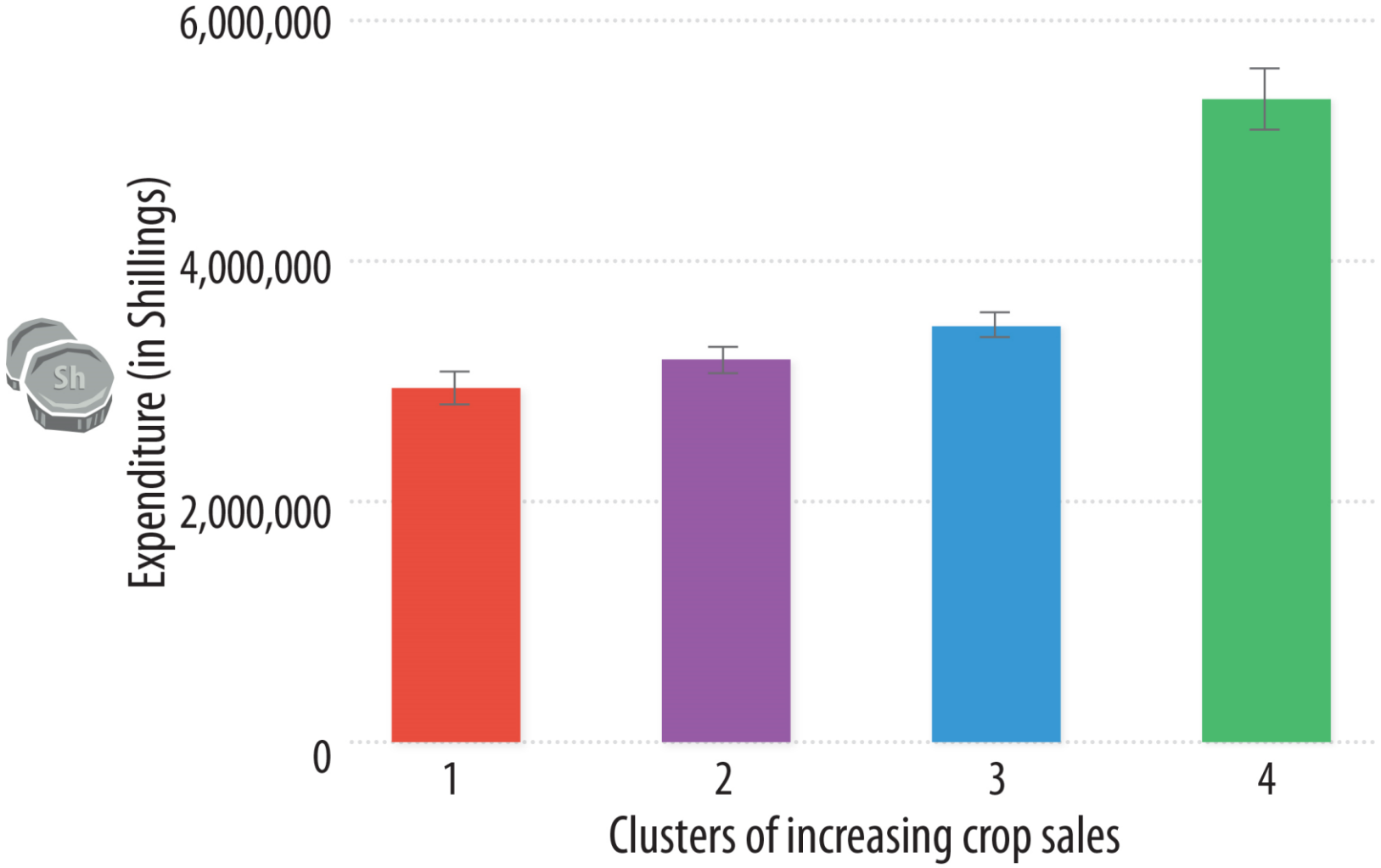
WORLD BANK GROUP

The variation in households without food deficiencies across clusters



WORLD BANK GROUP

The variation in expenditure across clusters



WORLD BANK GROUP

Optimizing income in a cluster

WORLD BANK GROUP

How to maximize income within a cluster

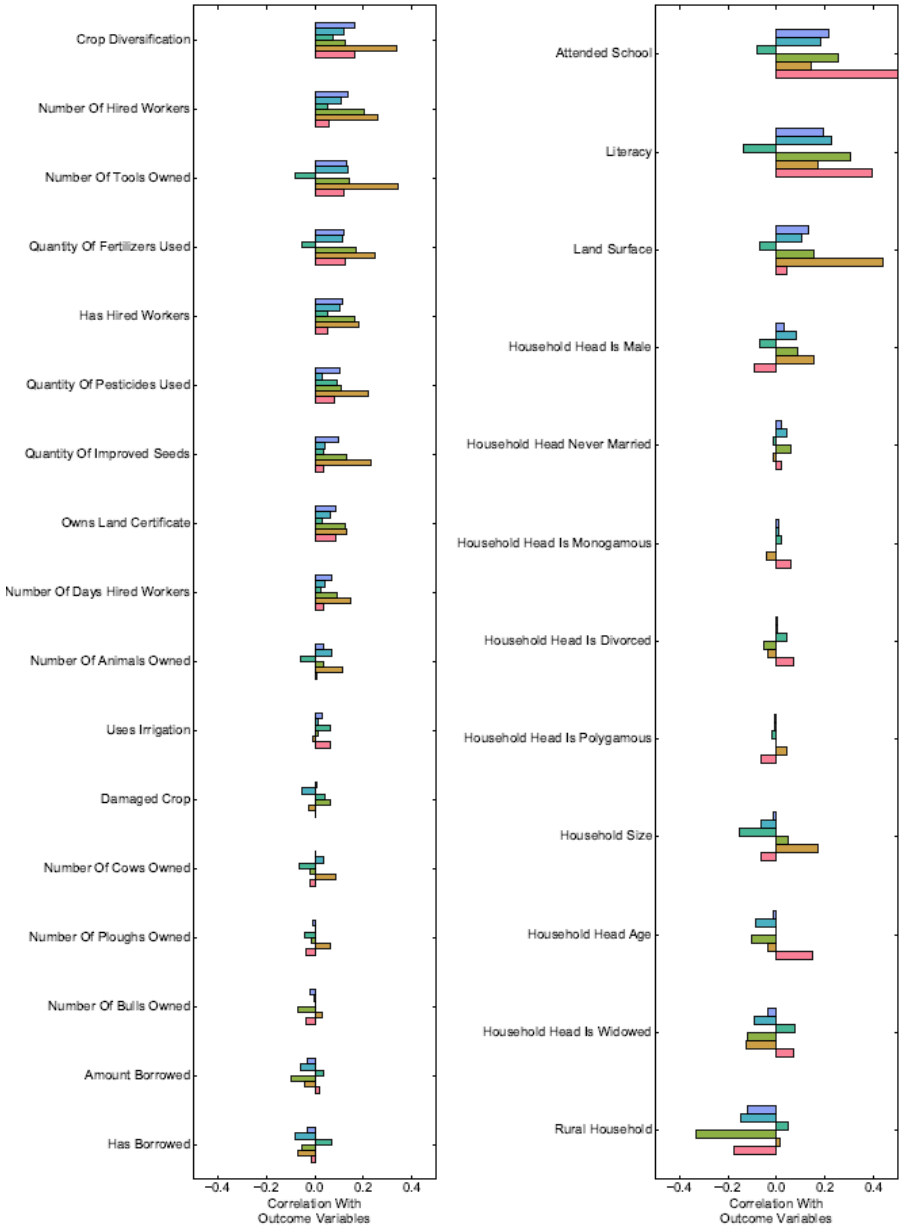
	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Most Impactful Input….	**Increase # of ploughs owned**	**Increase % who borrow money**	**Increase # of animals owned**	**Increase hiring of workers**
…which is highly correlated with these inputs	*No other variables*	*No other variables*	*No other variables*	*No other variables*
Input Coefficient	0.228	0.136	0.161	0.209
Input Coefficient Interpretation	Every 26% increase in % of farmers in cluster owns ploughs, is predicted to have94 485 shillings increase in average income in cluster	Every 47% increase in % who borrow money is predicted to have 203 928 shillings increase in average income in cluster	Every 46 additional animals owned, on average, is predicted to have 263298 shillings increase in average income in cluster	Every 25 additional workers hired is predicted to results in 474568 shillings increase in average income in cluster
Other Impactful Input….	**Increase hiring of workers**	**Increase # of animals owned**	**Increase # of ploughs owned**	**Increase irrigation use**
…. which is highly correlated with these inputs	*No other variables*	*Increase # of ploughs owned*	*Increase # of animals owned*	*No other variables*
Input Coefficient	0.21	0.131	0.118	0.156
Input Coefficient Interpretation	Every 11 additional workers hired is predicted to result in 87 025 shillings increase in average income in cluster	Every 36 additional animals owned, on average, is predicted to result in 196 423 shillings increase in average income in cluster	Every 62% average increase in proportion of farmers who own ploughs, is predicted to result in 192 967 shillings increase in average income in cluster	Every 15% increase in irrigation usage within cluster, is predicted to result in 354 223 shillings increase in average income in cluster

UGANDA: Results Overview

WORLD BANK GROUP

Correlation between Uganda inputs and outcomes

- Correlations between inputs and outcomes are of lower order of magnitude as Ethiopia's correlations, and similar to Tanzania's
- Strength of correlations for individual inputs are not the same for the 3 countries: **points to possible different pathways and optimized input variables**



Actionable inputs Non-actionable inputs

WORLD BANK GROUP

List of most correlated inputs for k-means clustering of households in Uganda's LSMS-ISA data

- Whether the farmer has hired workers (but not the number of workers hired)
- The number of animals owned
- Quantity of pesticides used
- Crop diversification
- Household size
- Whether the household head is separated

How are the clusters different? *(actionable variables)*

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Amount Borrowed	1388.501	706.973	1074.525	2566.904
Number of Crops Planted	4.214	10.716	12.359	15.595
% of Households Who Borrowed Money	0.271	0.337	0.25	0.215
Has Hired Workers	0.483	0.464	0.564	0.694
Number Of Days Hired Workers	4.29	4.184	7.38	13.839
Number Of Hired Workers	12.285	10.378	18.646	42.467
Number Of Animals Owned	1.137	0.96	1.321	3.842
Number Of Bulls Owned	0.158	0.097	0.211	0.492
Number Of Cows Owned	0.566	0.517	0.66	2.271
Number Of Ploughs Owned	0.063	0.015	0.09	0.141
Number Of Tools Owned	3.968	4.566	6.239	13.39
% of Households Who Own Land Certificate	0.439	0.589	0.694	1.171
Quantity Of Fertilizers Used (in Kg)	36.891	29.53	161.588	515.747
Quantity Of Improved Seeds Used (in Kg)	159.752	196.421	270.545	479.099
Quantity Of Pesticides Used (in Liters)	61.613	80.385	84.693	511.852
% of Households Using Irrigation	0.039	0.096	0.075	0.076

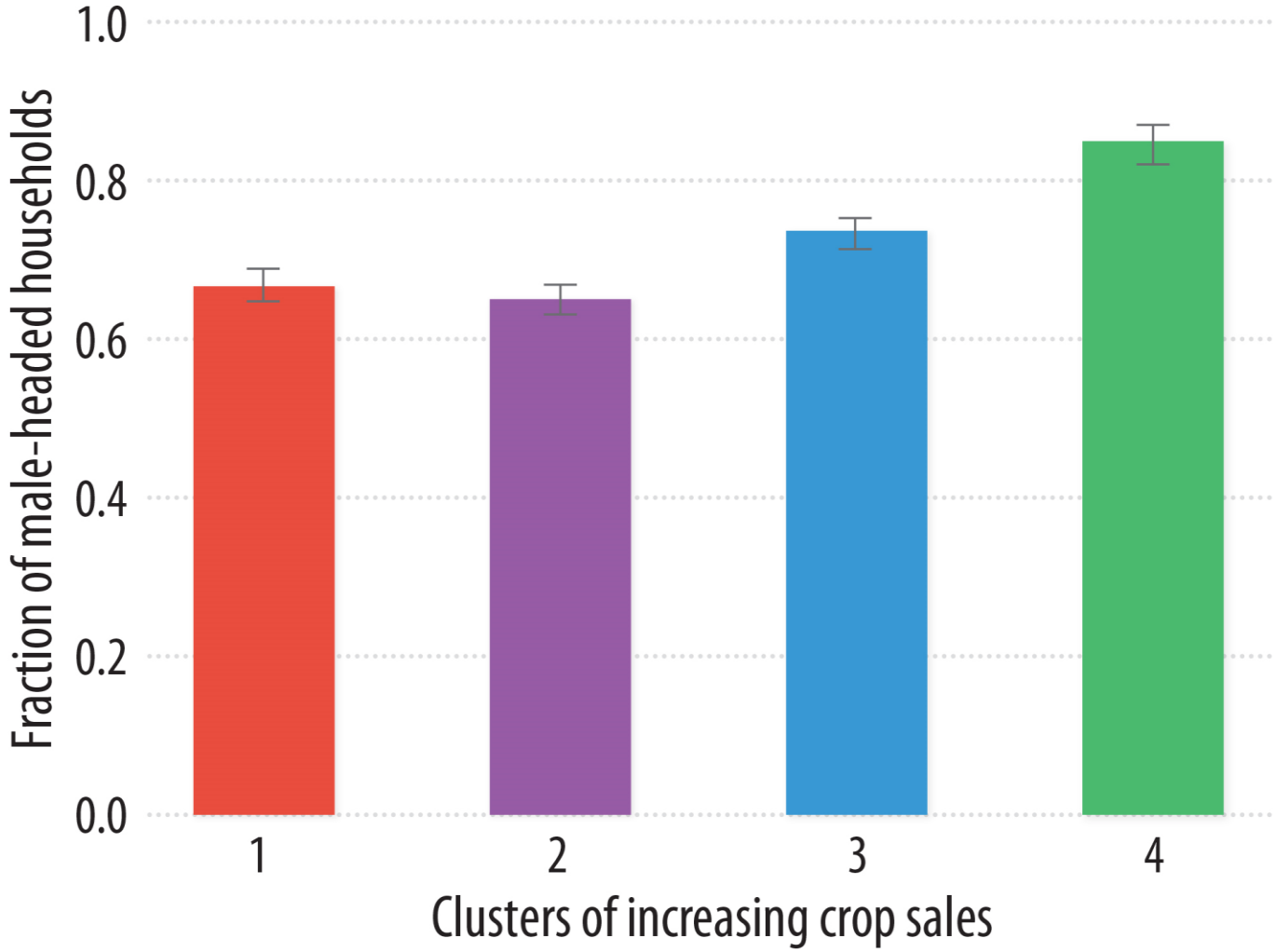
How are the clusters different? (non-actionable variables and outcomes)

NON ACTIONABLE	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Fraction of Households Attending School	0.691	0.747	0.786	0.864
Household Head Age	44.132	46.27	47.399	50.68
Fraction of Divorced	0.086	0.085	0.056	0.025
Fraction of Male-headed Households	0.667	0.649	0.732	0.845
Household Head Is Monogamous	0.838	0.815	0.788	0.776
Household Head Is Polygamous	0.162	0.185	0.212	0.224
Fraction of Widow	0.152	0.18	0.127	0.07
Household Head Never Married	0.026	0.033	0.009	0.006
Number of Household Members	5.174	5.145	6.513	8.543
Land Surface (in Ha)	1.684	1.723	5.06	9.808
Literacy Rate	0.519	0.56	0.618	0.727
Rural Household	0.785	0.888	0.915	0.871
OUTCOMES	**CLUSTER 1**	**CLUSTER 2**	**CLUSTER 3**	**CLUSTER 4**
Children Education	0.727	0.779	0.806	0.843
Crop Sales (in Shilling)	260131.5	371814.5	1075892	2516626
Expenditure (in Shilling)	363122.9	357314.5	380455.4	537182.8
Medical Assistance	0.302	0.336	0.28	0.214
% of Households Without Food Deficiencies	0.691	0.759	0.791	0.892

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

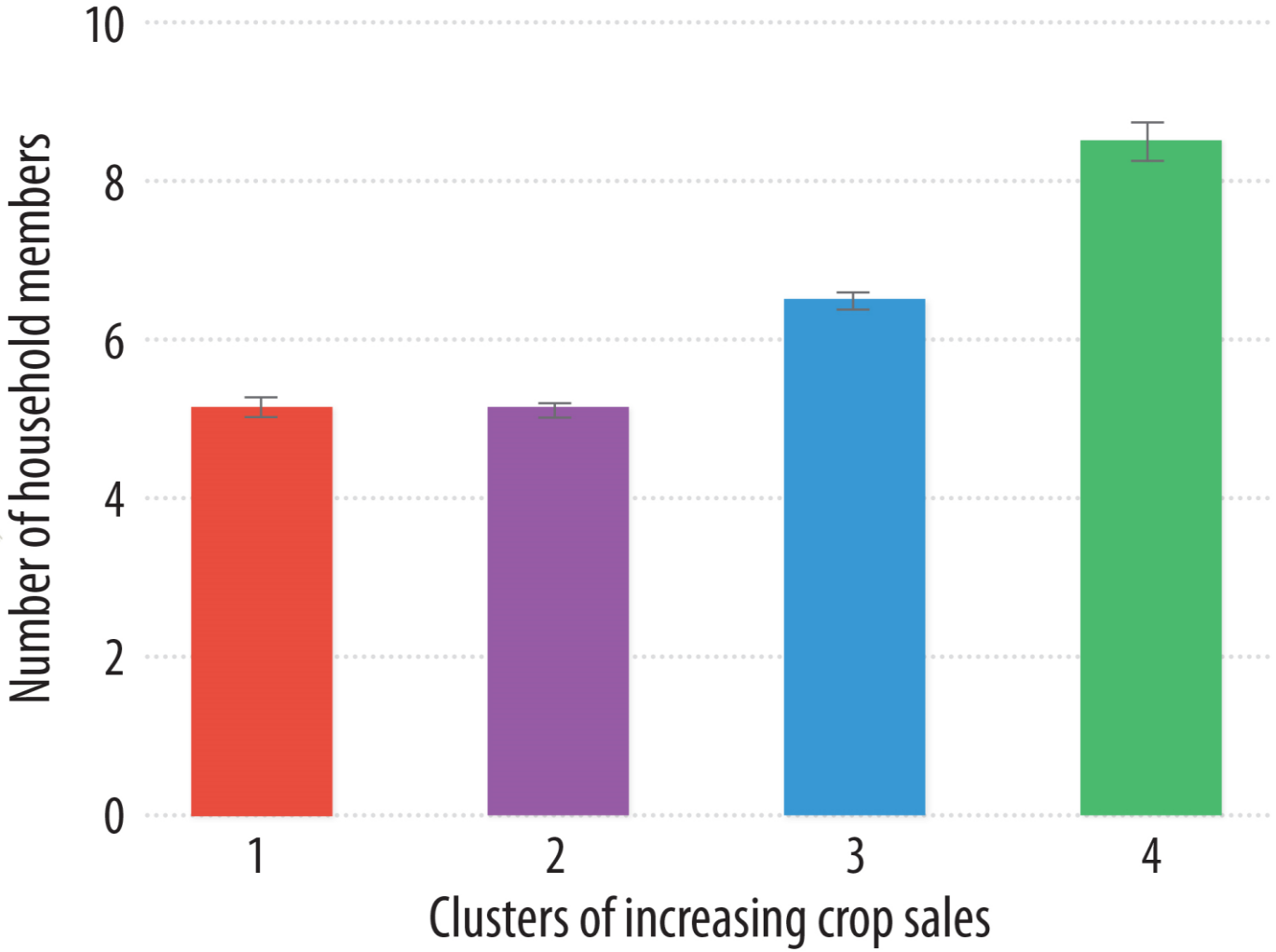
First, variation in non-actionable variables

The variation in male-headed household across clusters



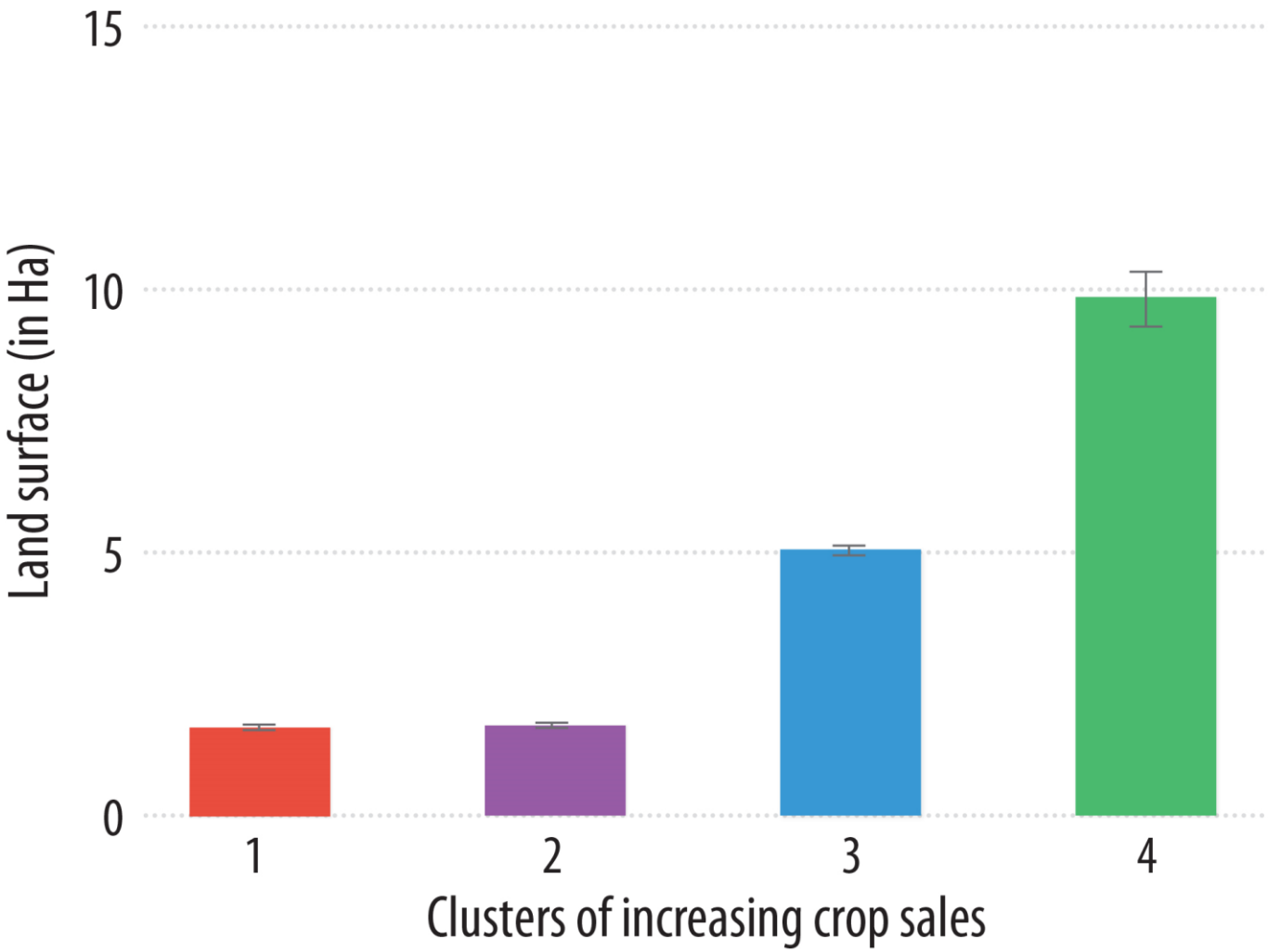
WORLD BANK GROUP

The variation in number of household members across clusters



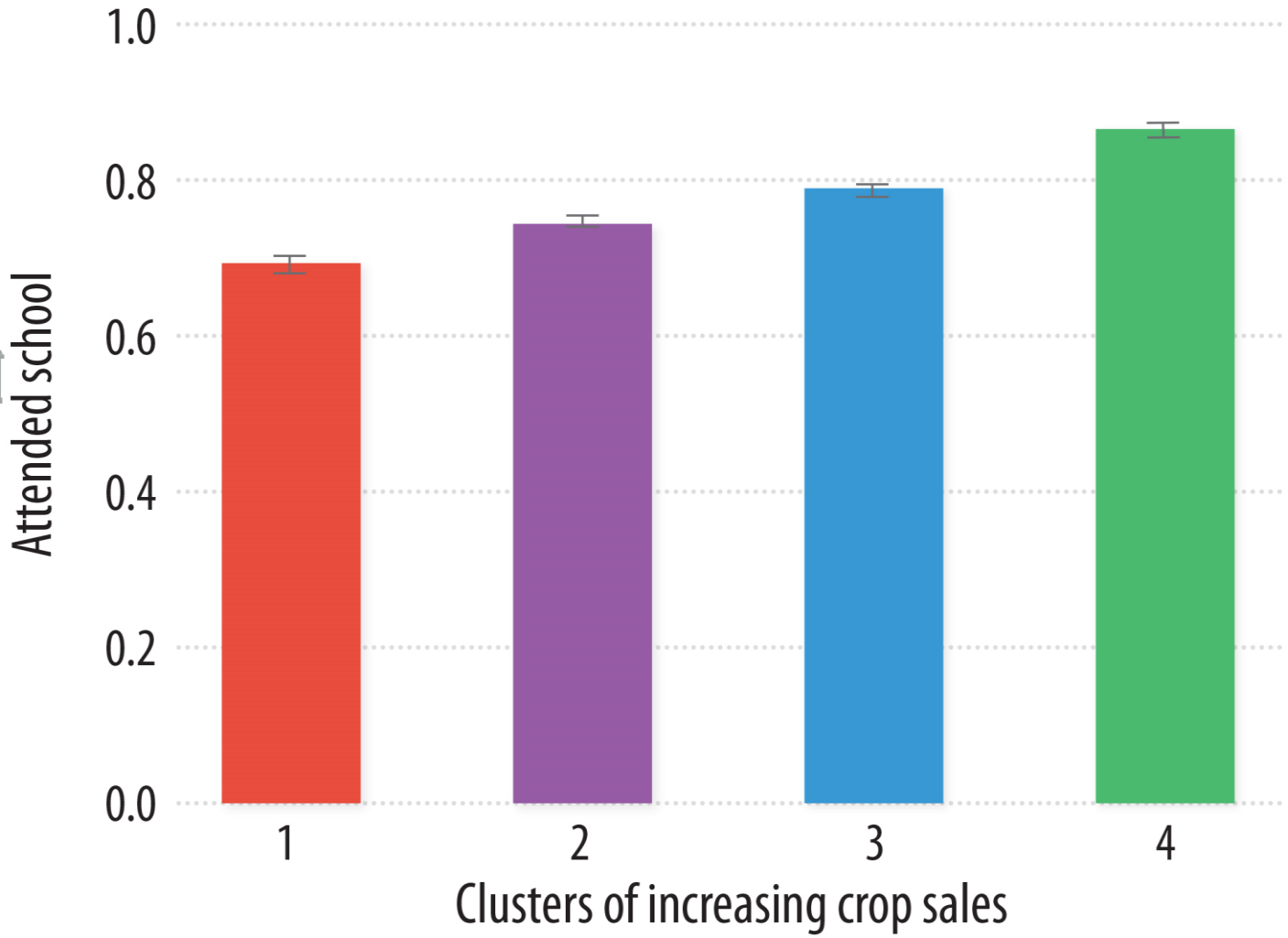
WORLD BANK GROUP

The variation in land surface across clusters



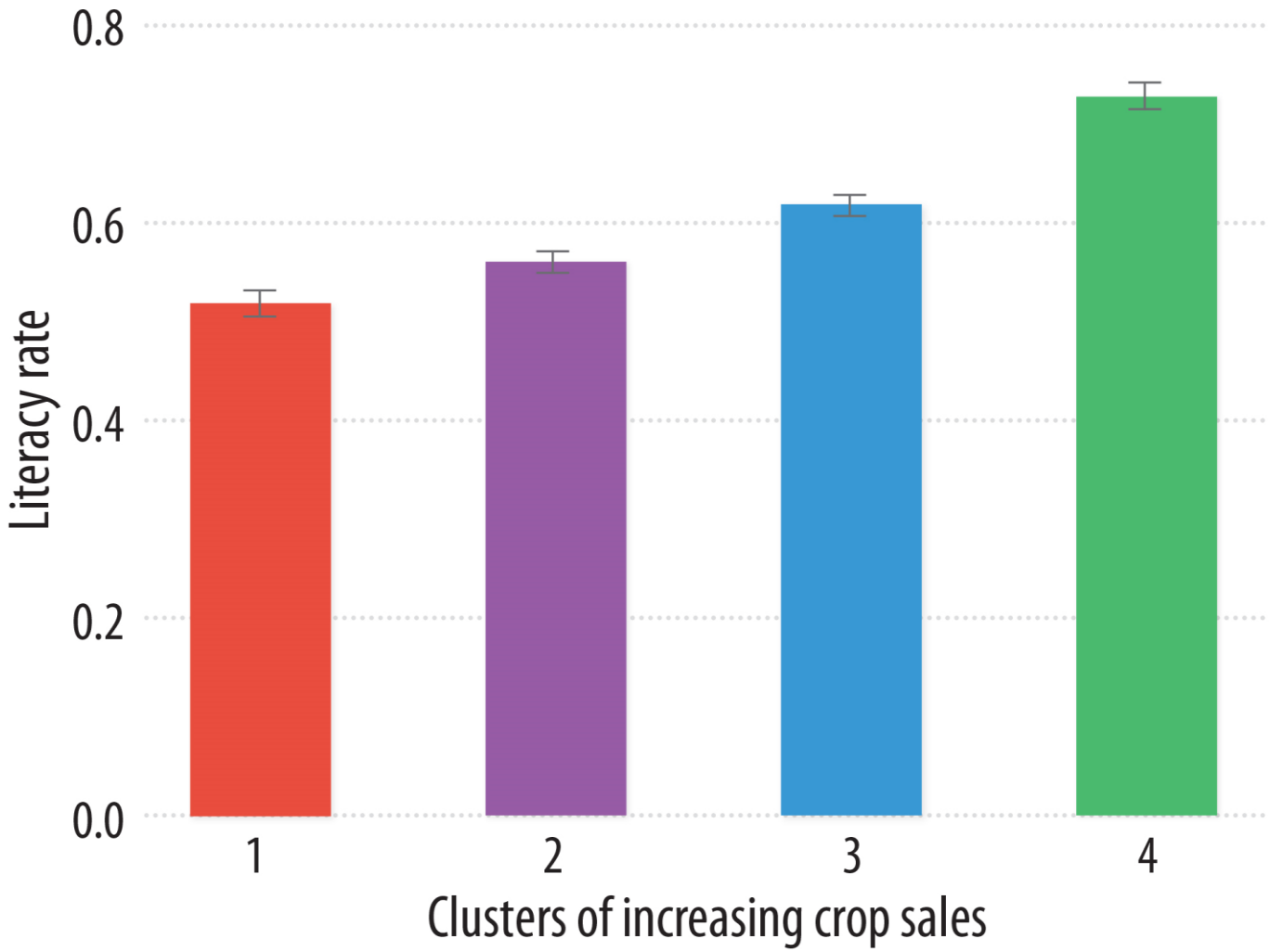
WORLD BANK GROUP

The variation in attended school across clusters



WORLD BANK GROUP

The variation in literacy rate across clusters

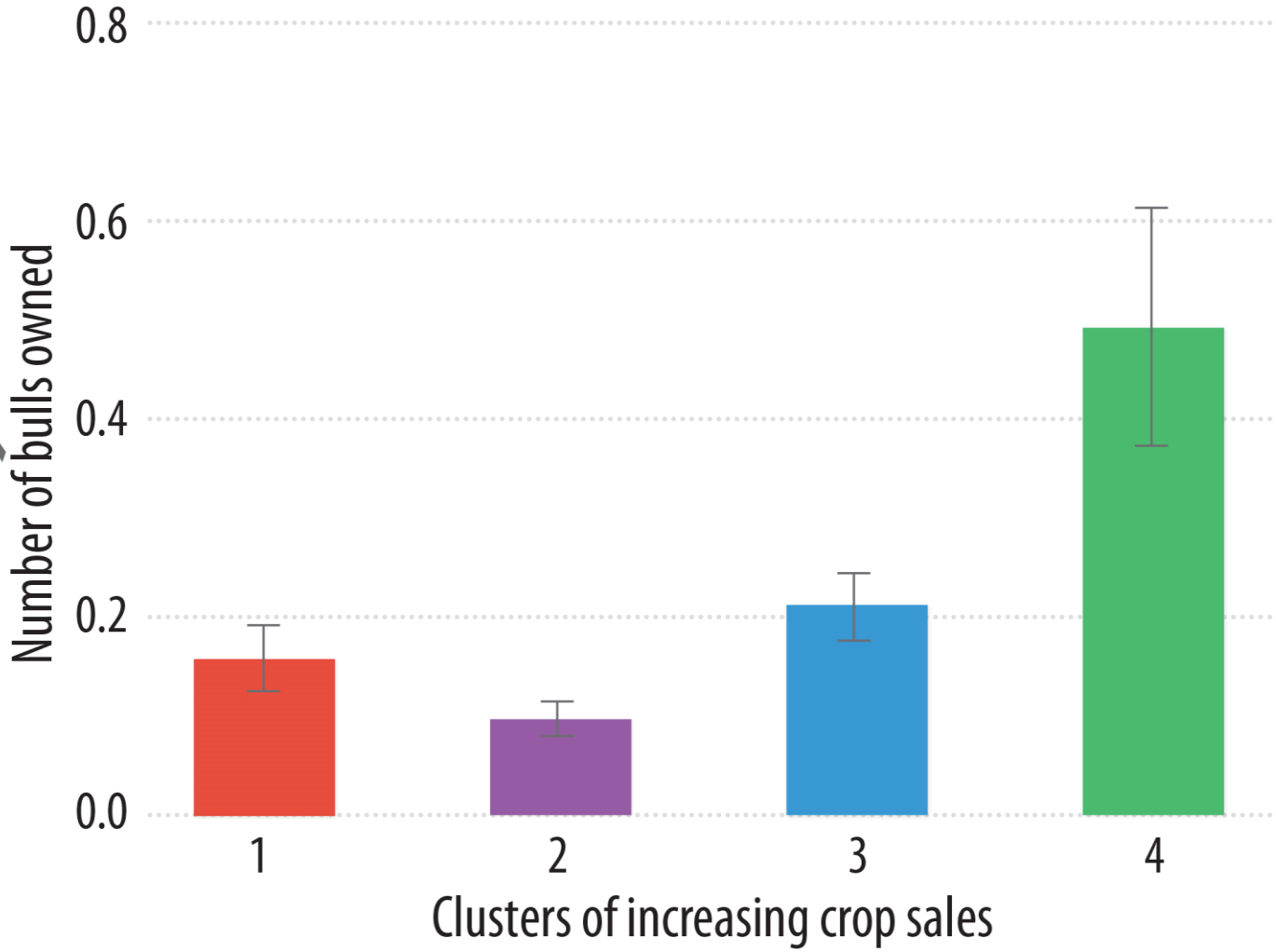


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

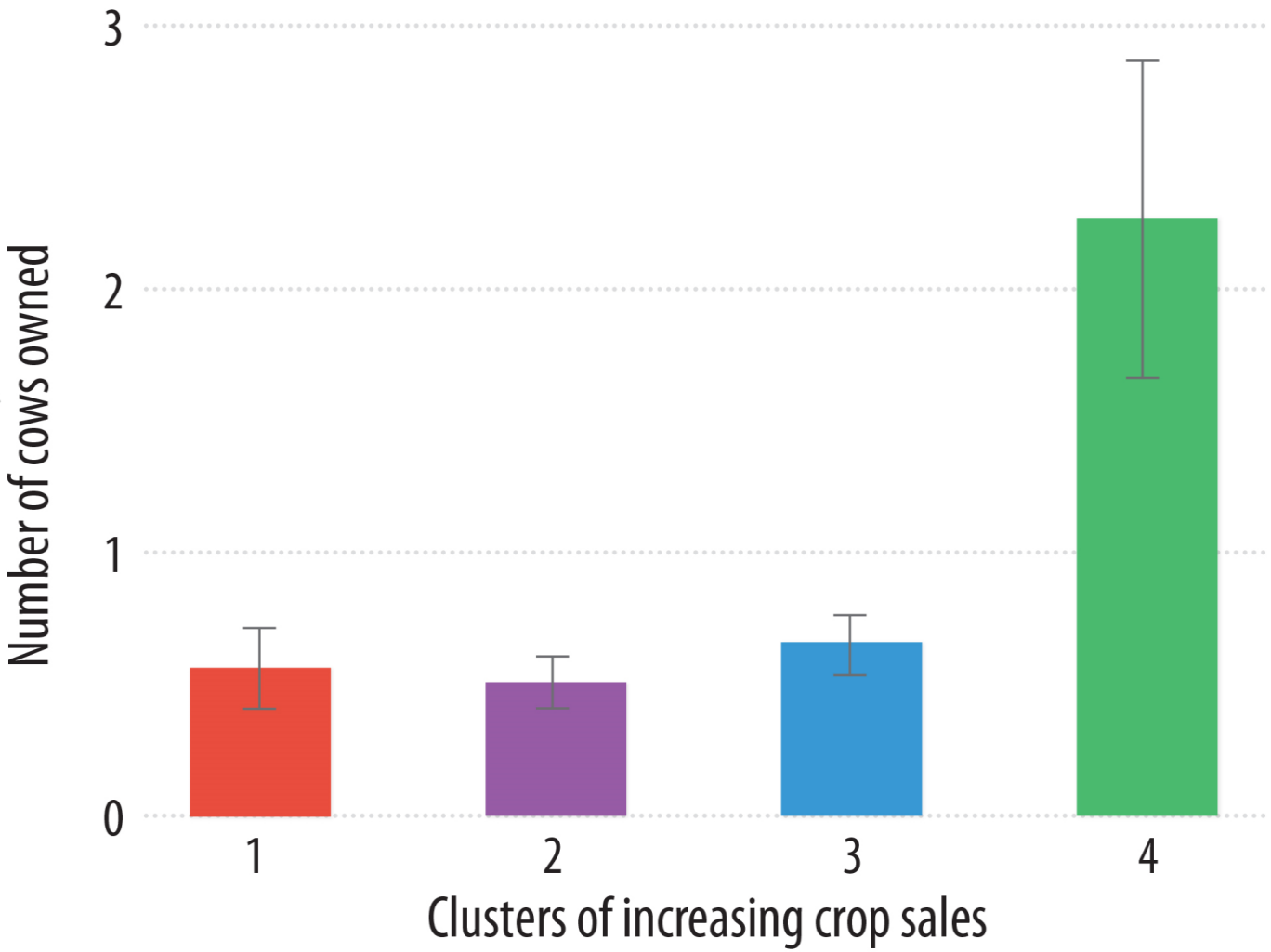
Second, variation in actionable variables

The variation in number of bulls owned across clusters



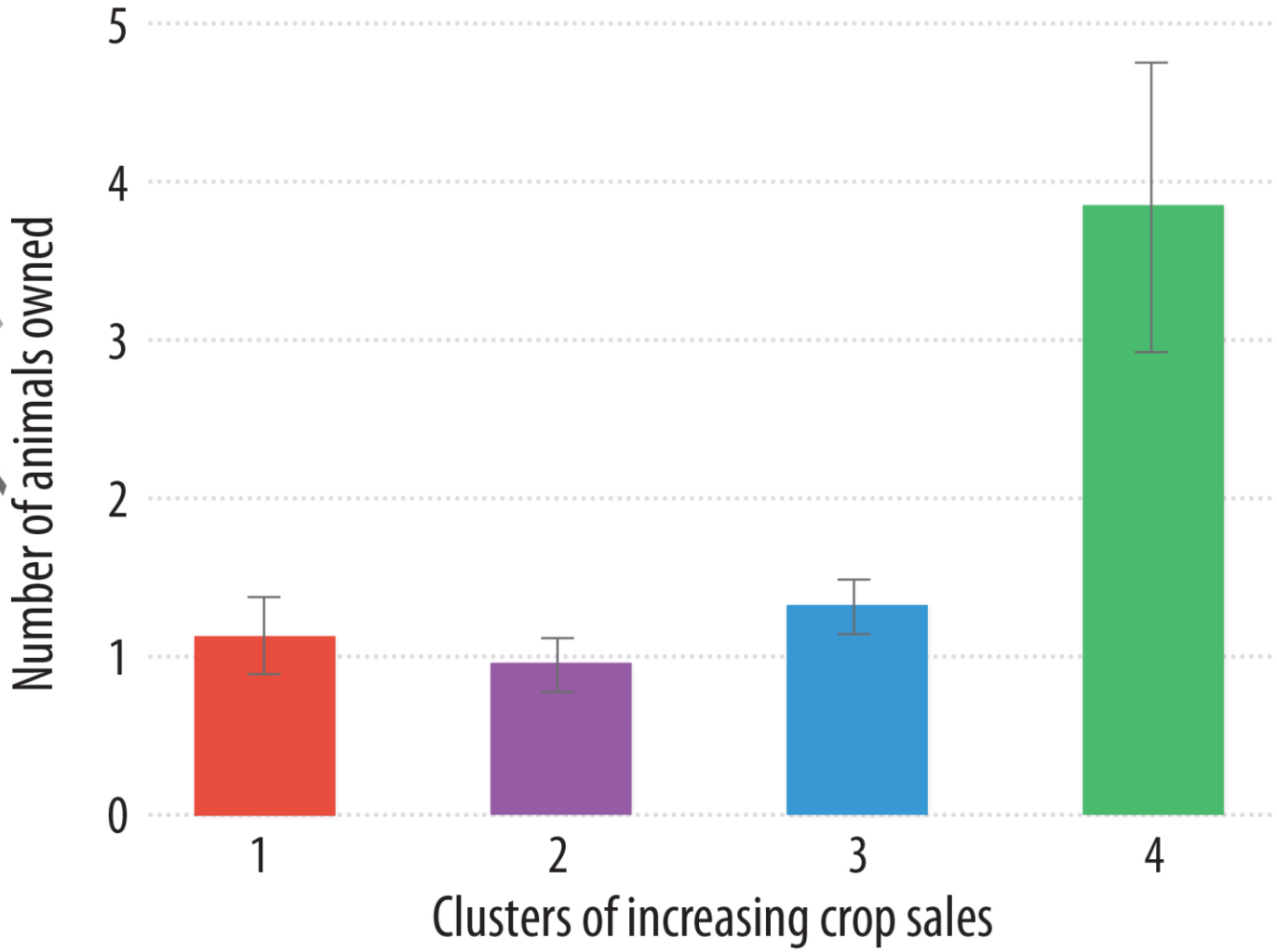
WORLD BANK GROUP

The variation in number of cows owned across clusters



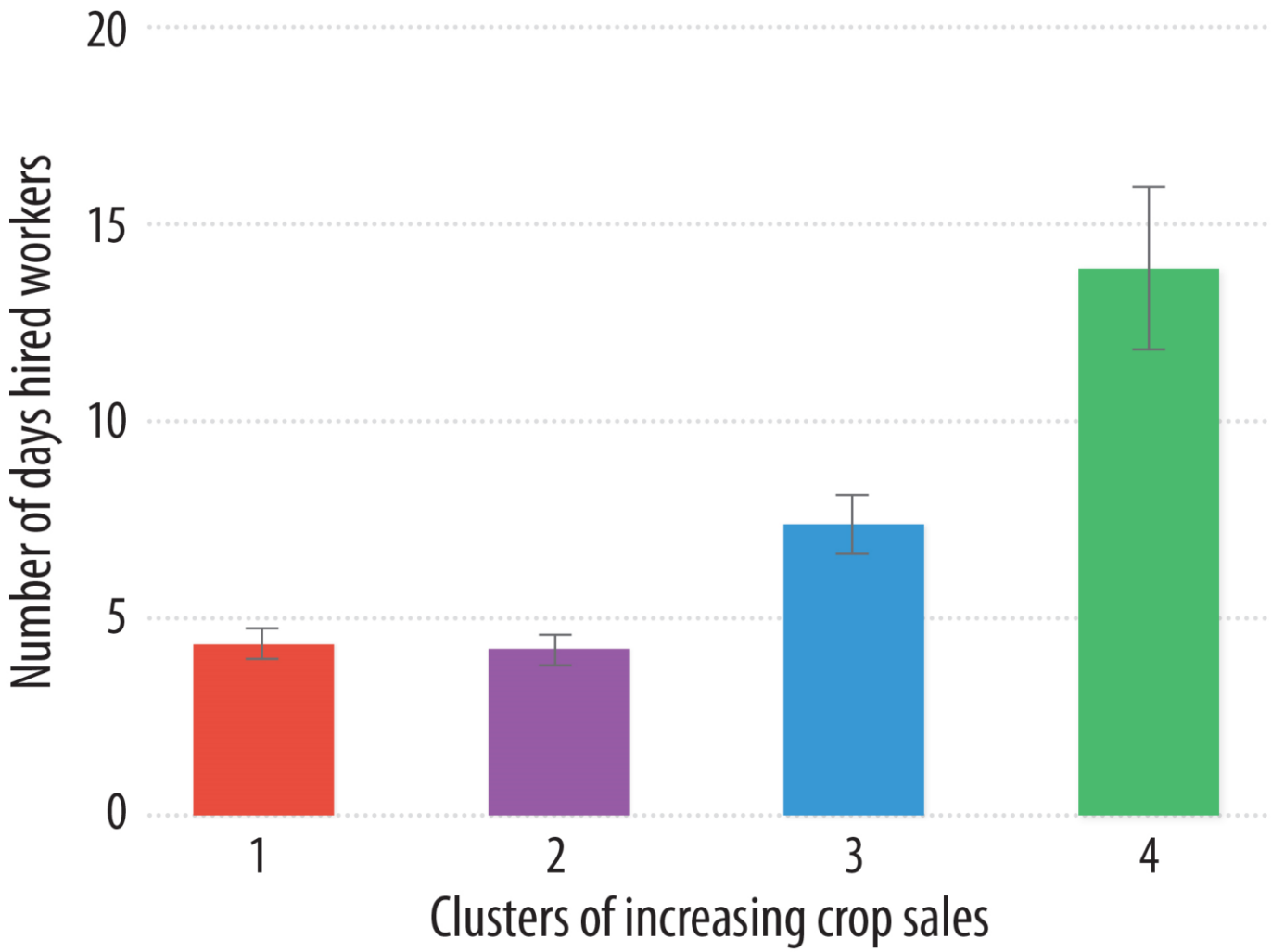
WORLD BANK GROUP

The variation in number of animals owned across clusters



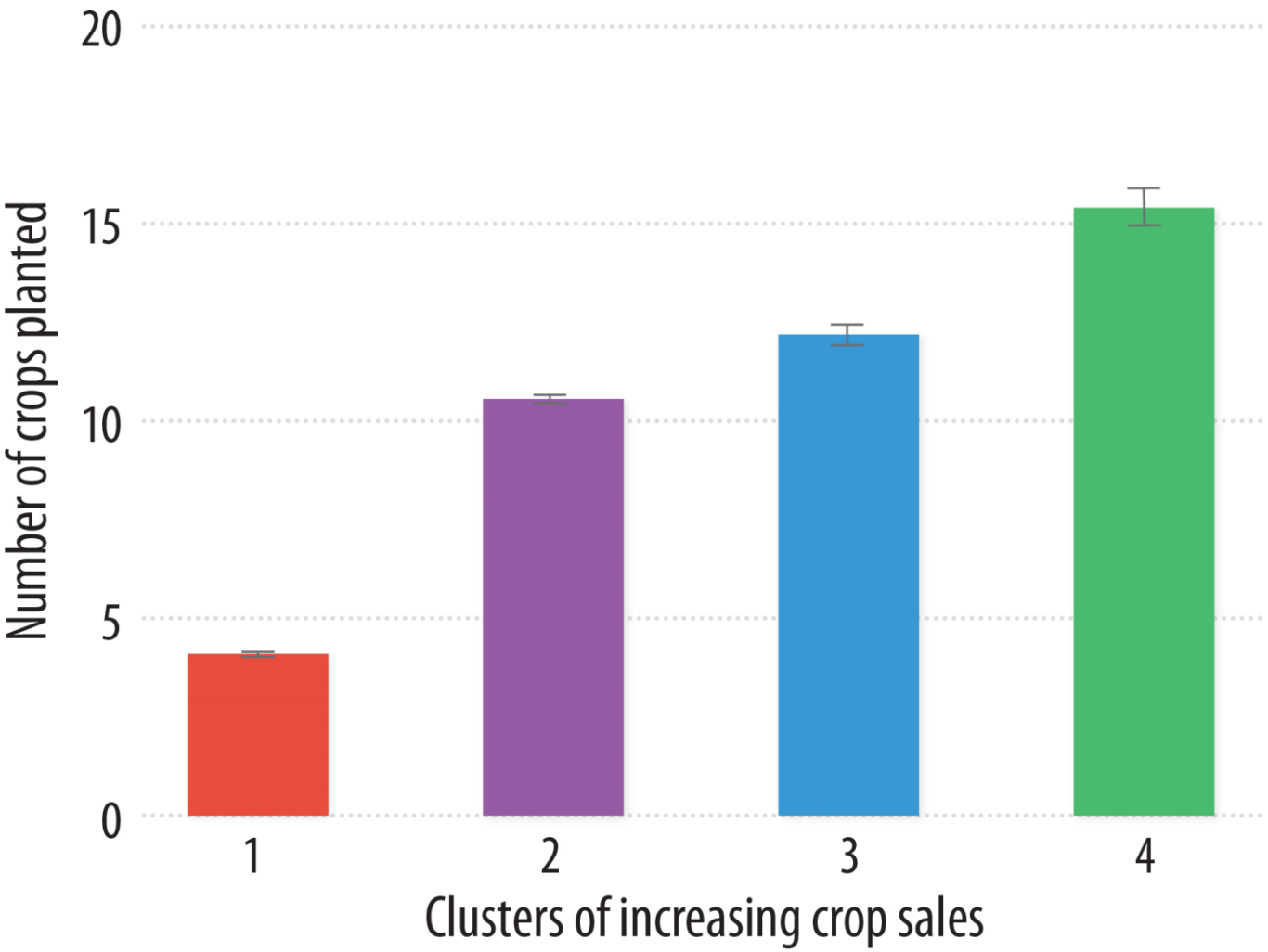
WORLD BANK GROUP

The variation in number of days hired workers across clusters



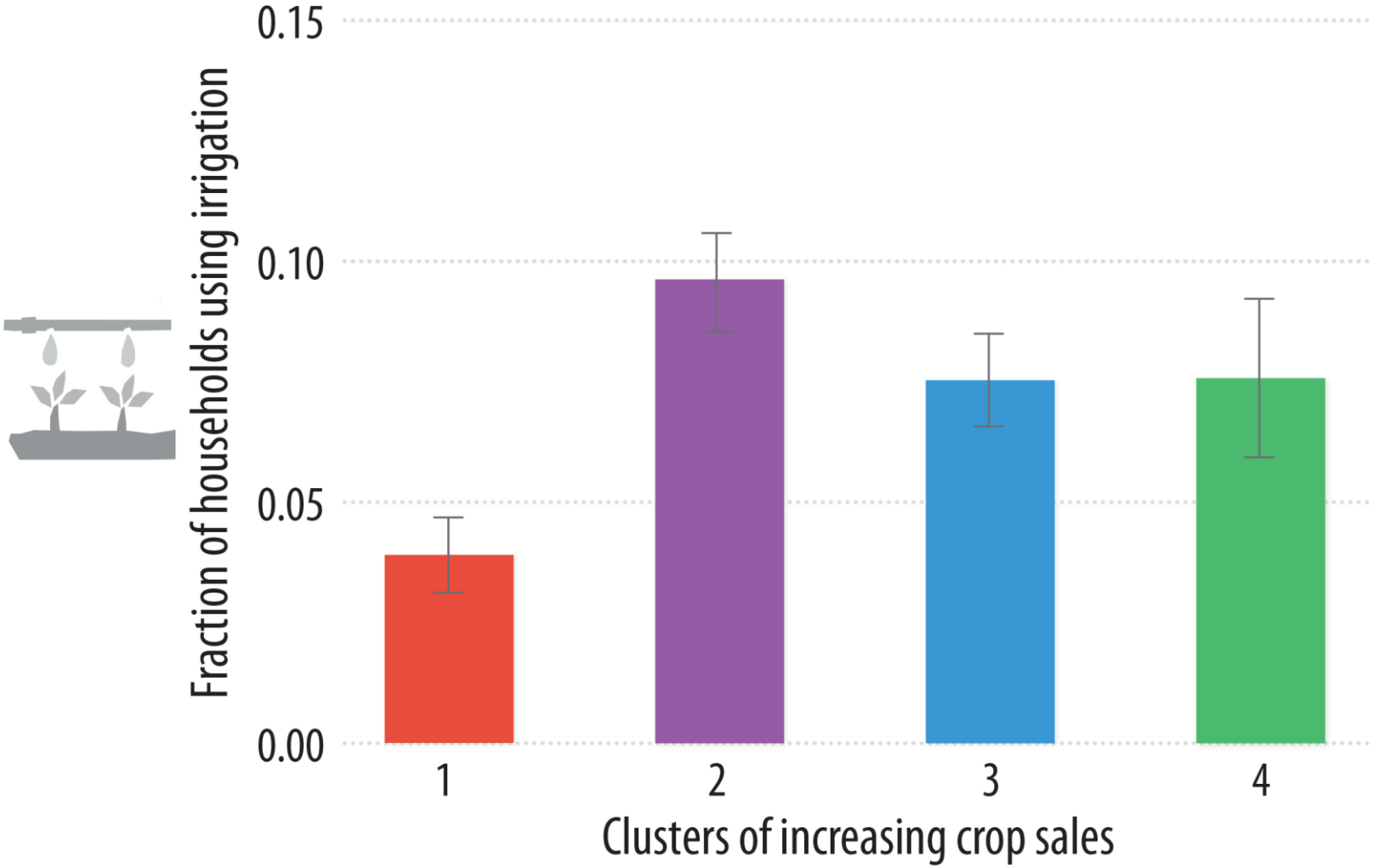
WORLD BANK GROUP

The variation in crop diversification across clusters



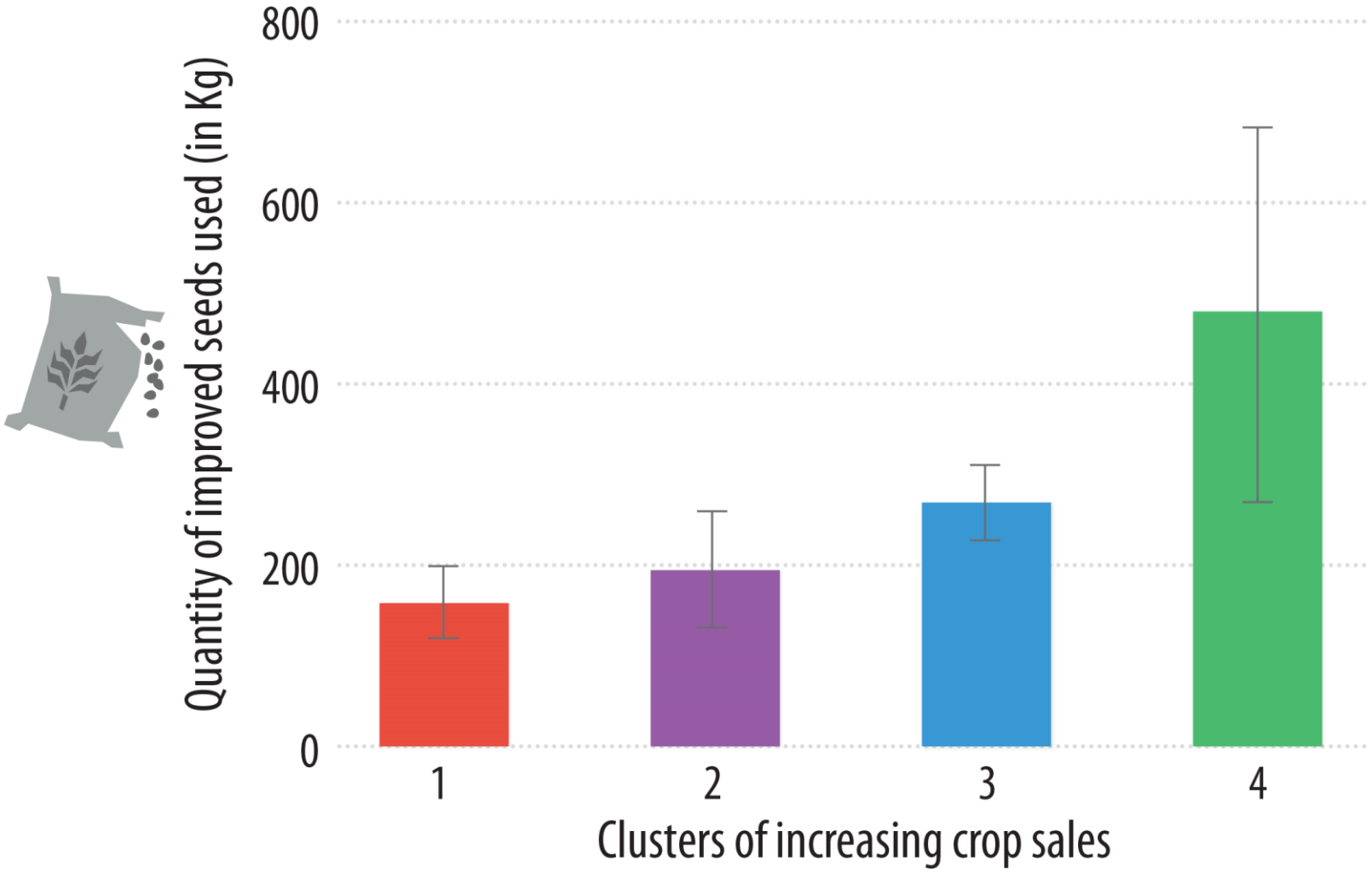
WORLD BANK GROUP

The variation in households using irrigation across clusters



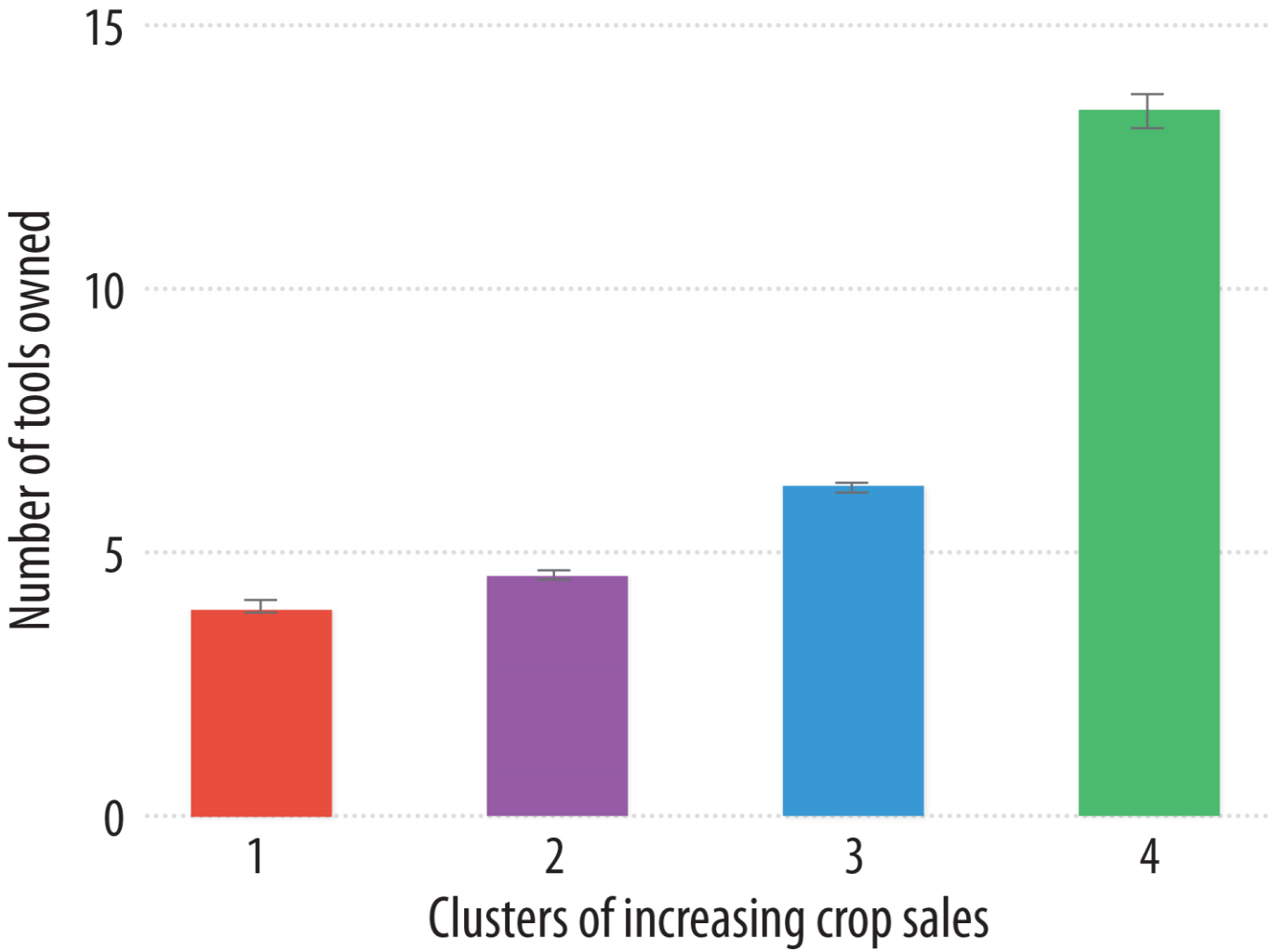
WORLD BANK GROUP

The variation in quantity of improved seeds used across clusters



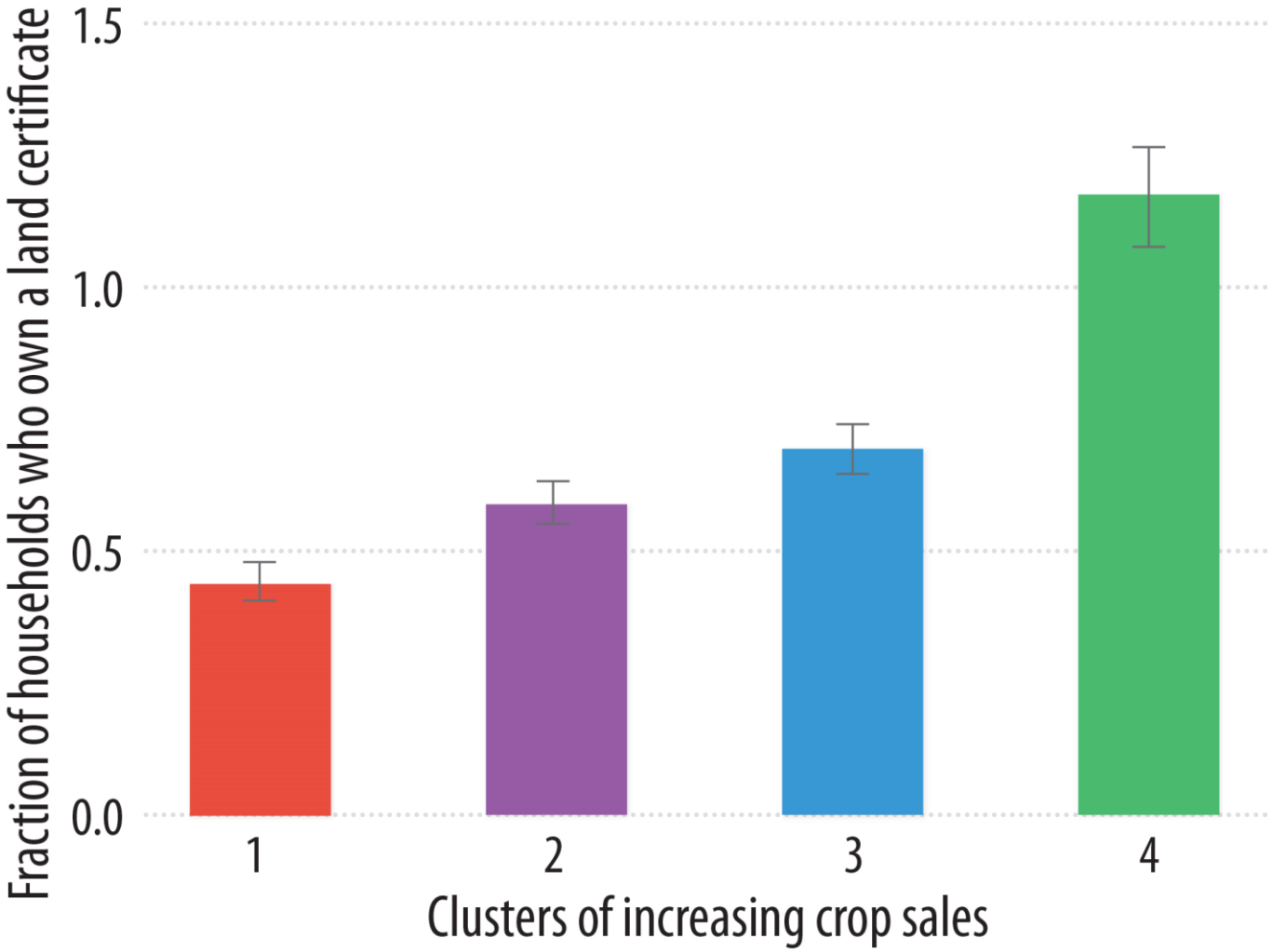
WORLD BANK GROUP

The variation in number of tools owned across clusters



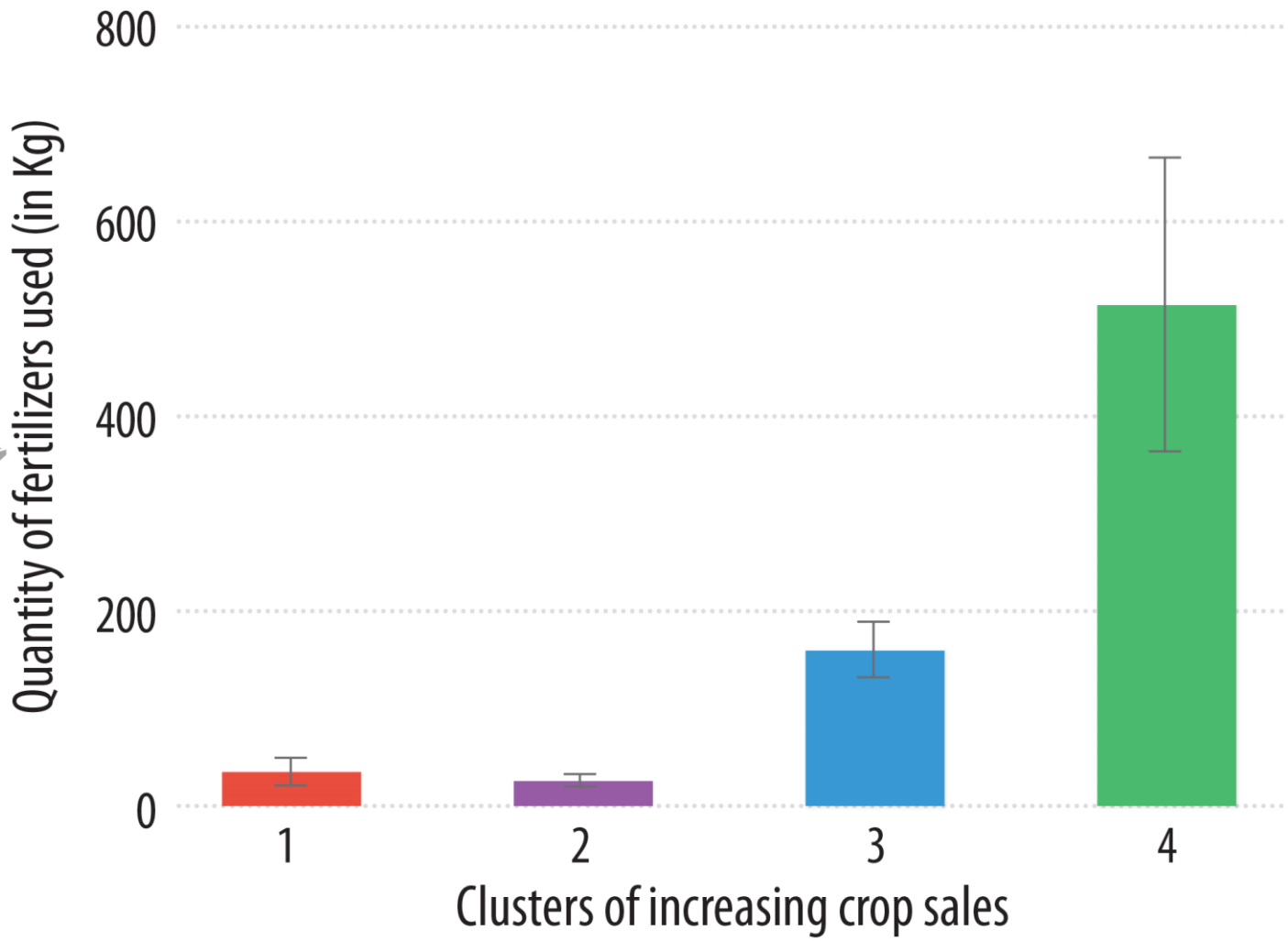
WORLD BANK GROUP

The variation in % of households who own a land certificate across clusters



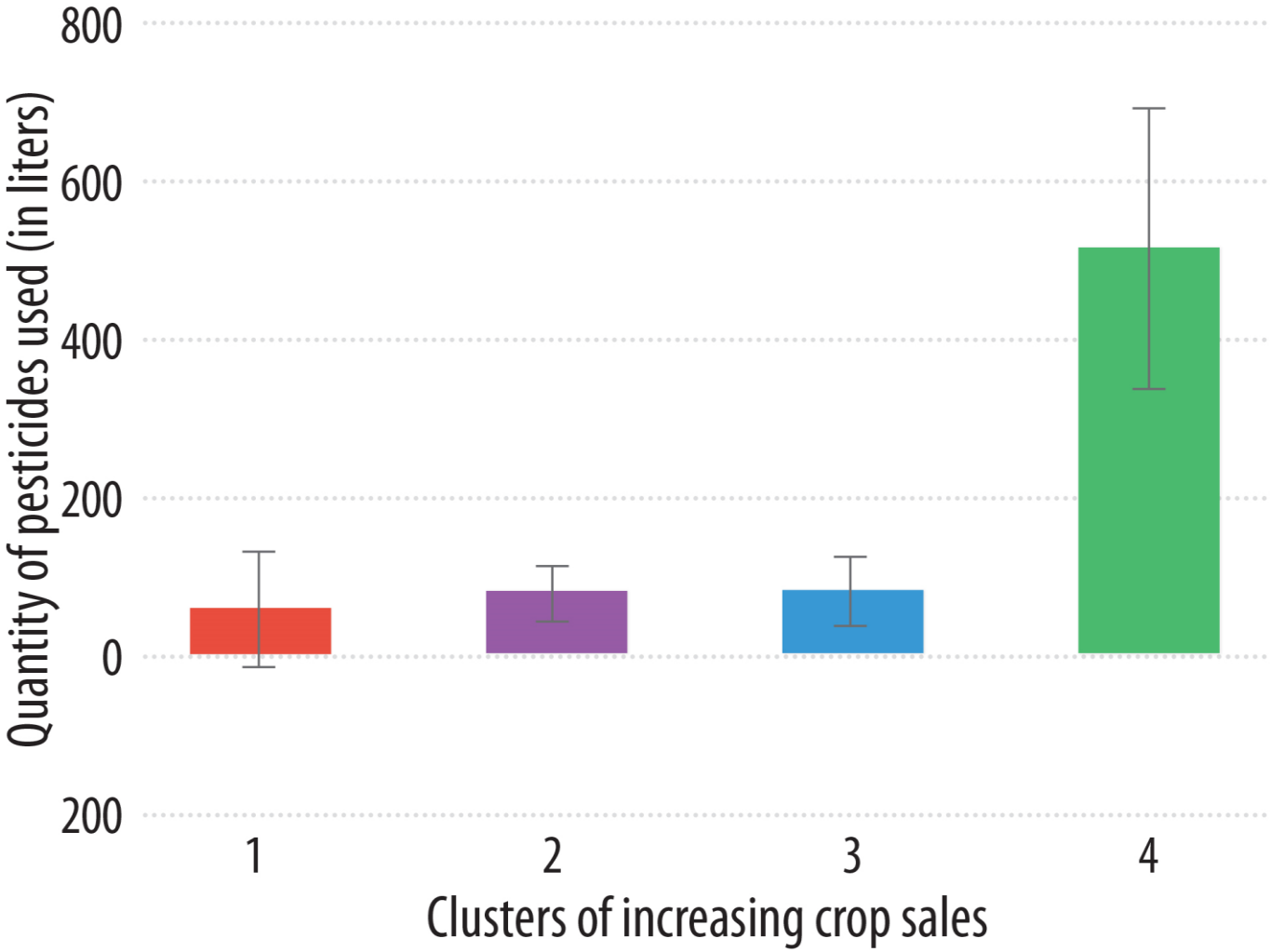
WORLD BANK GROUP

The variation in quantity of fertilizers used across clusters



WORLD BANK GROUP

The variation in quantity of pesticides used across clusters

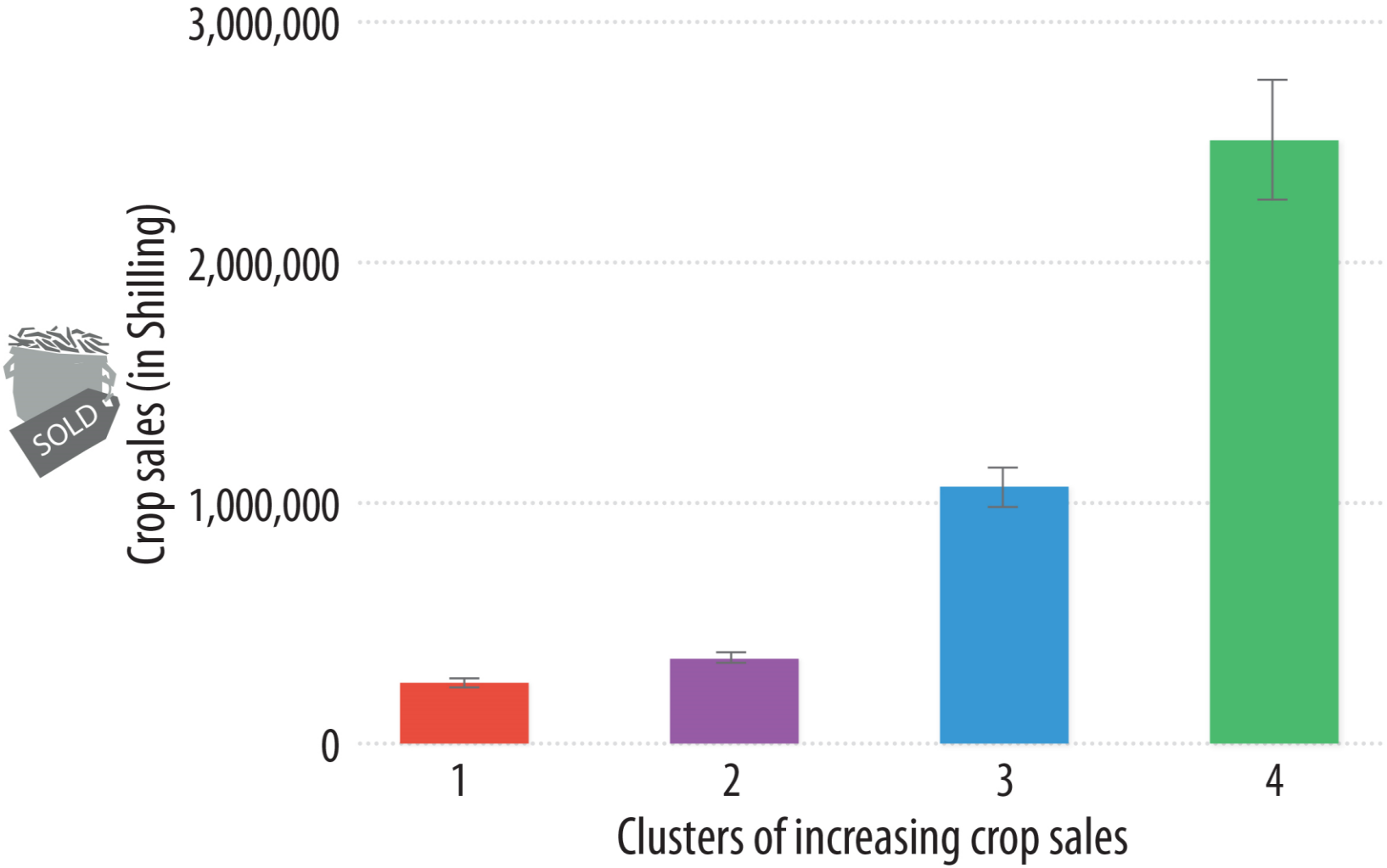


WORLD BANK GROUP

Now that we **have clustered farmers** and have a better sense of what each cluster of farmers look like, we look at variation of identified variables across clusters in order to make policy recommendations.

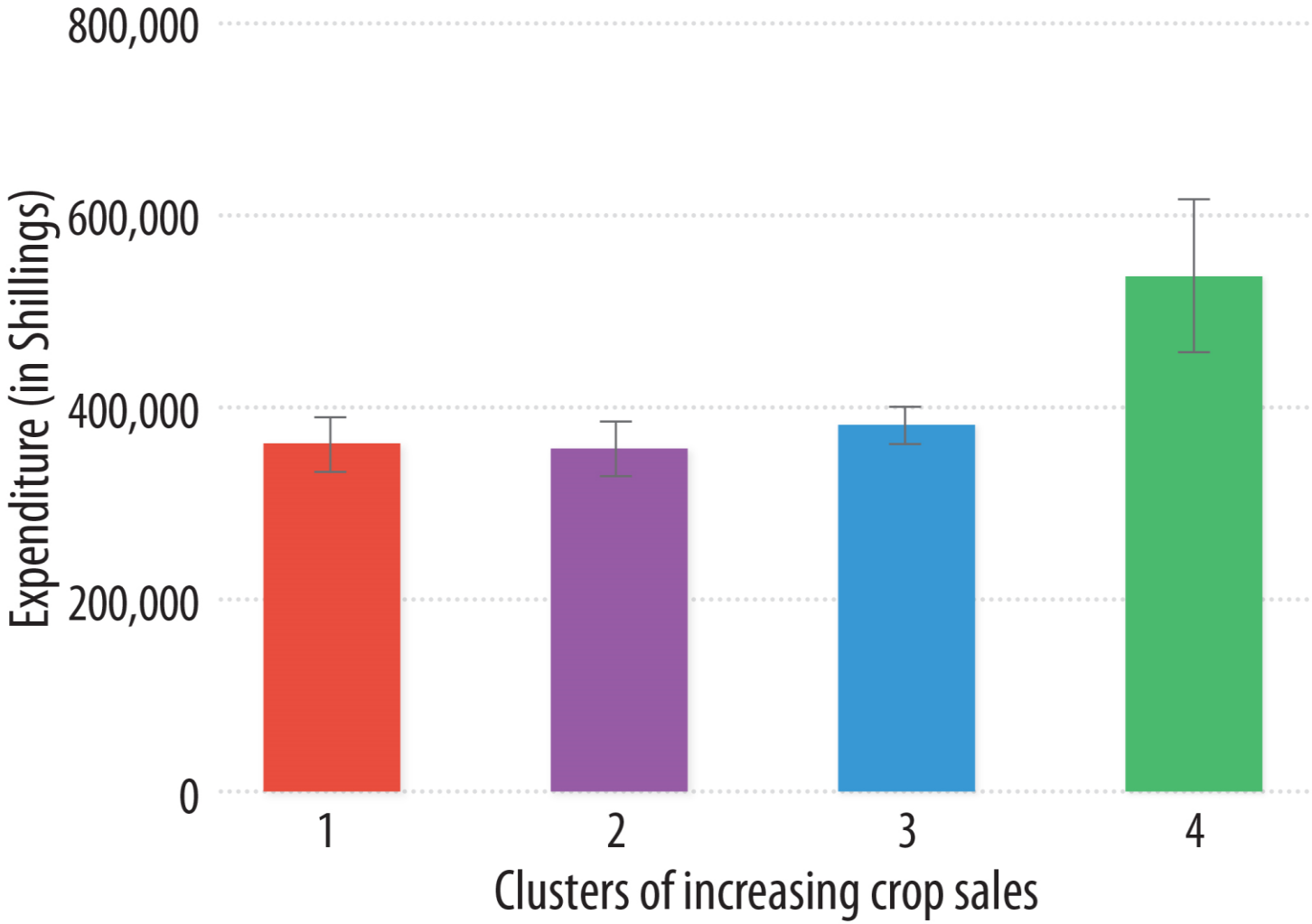
Third, variation in outcomes

The variation in crop sales across clusters



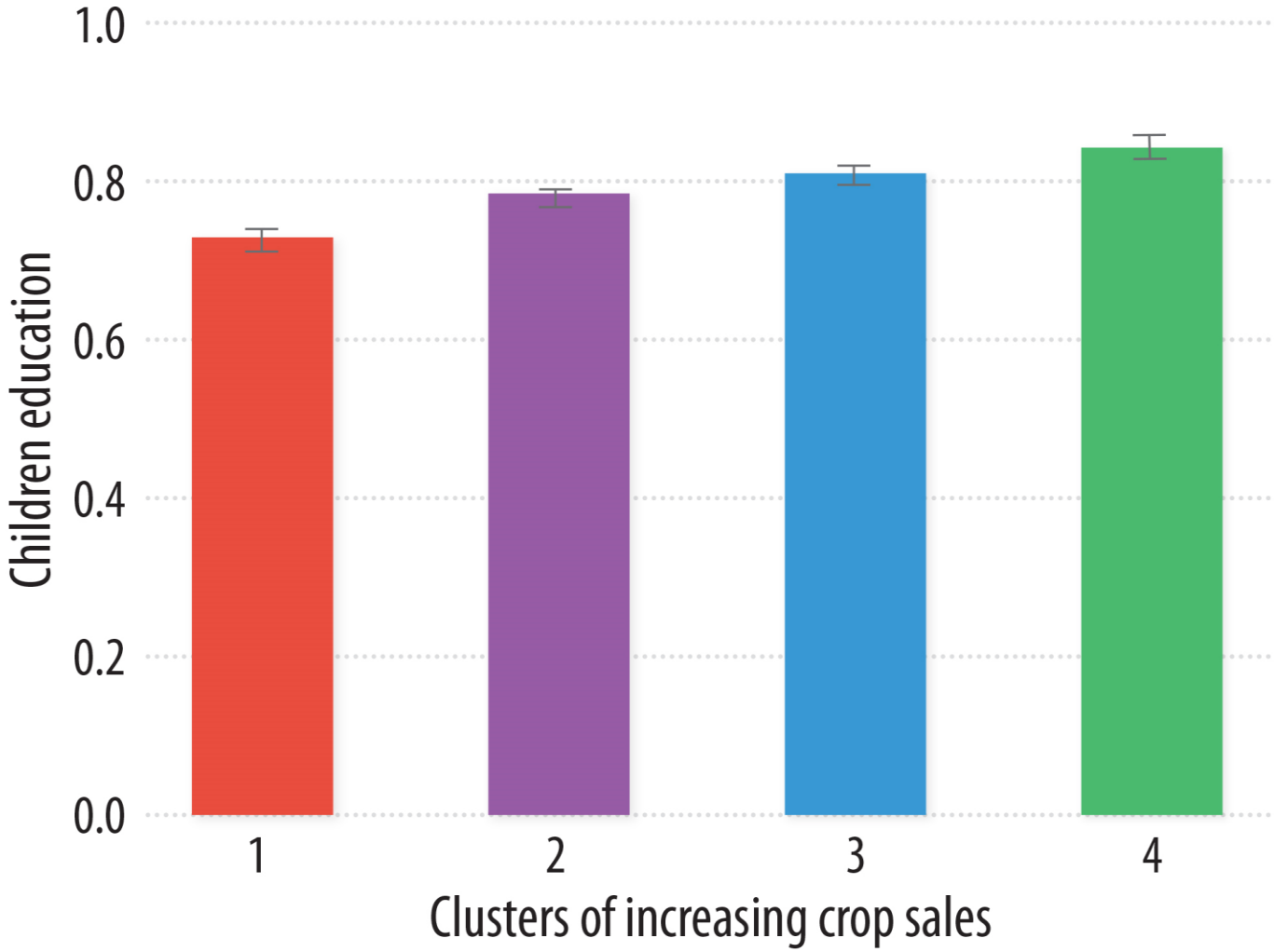
WORLD BANK GROUP

The variation in expenditure across clusters

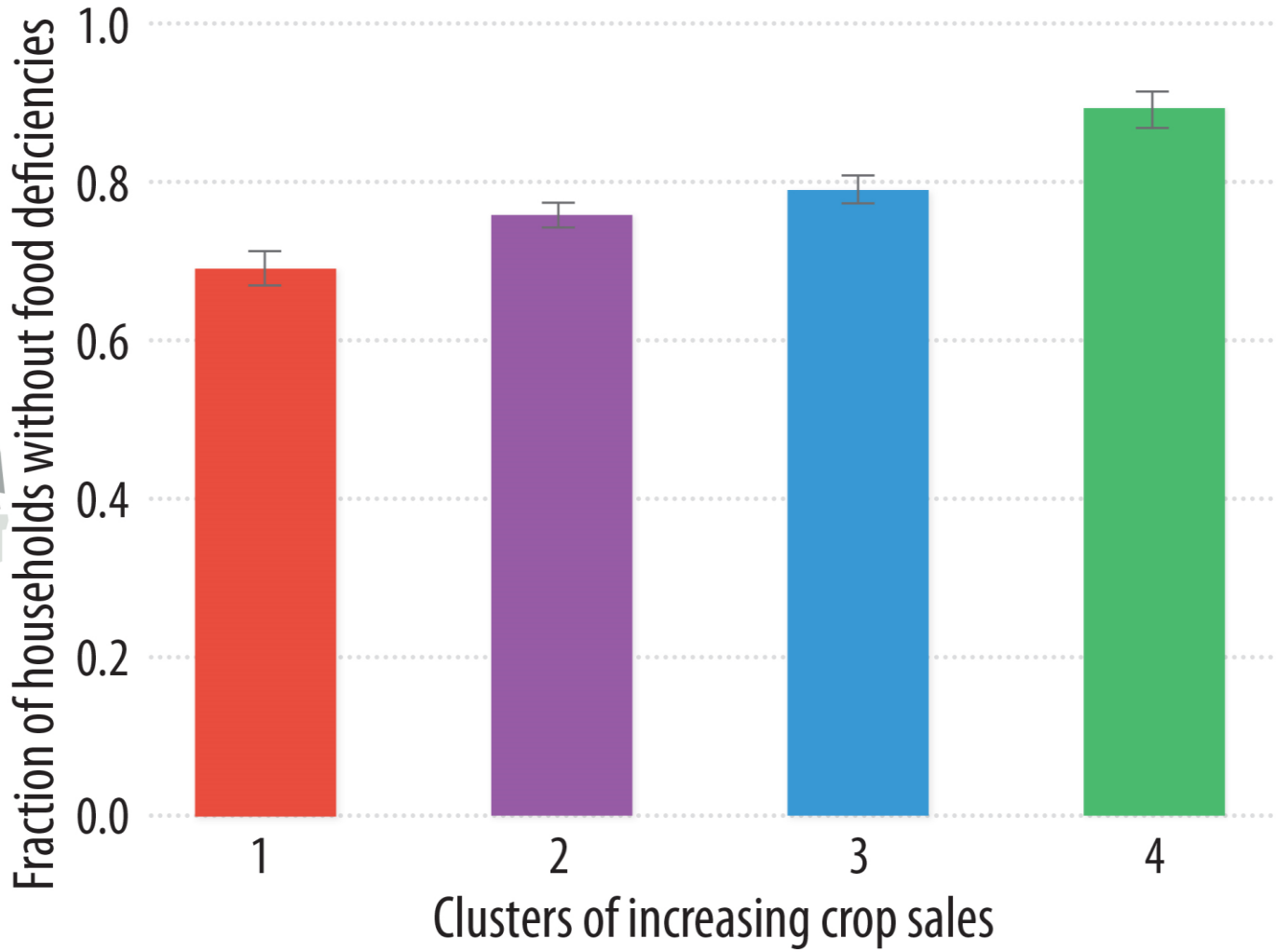


WORLD BANK GROUP

The variation in children education across clusters



The variation in households without food deficiency across clusters



WORLD BANK GROUP

Optimizing income in a cluster

WORLD BANK GROUP

How to maximize income within a cluster

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Most Impactful Input….	**Increase number of days for which workers are hired**	**Decrease crop diversification**	**Increase number of days for which workers are hired**	**Decrease crop diversification**
…which is highly correlated with these inputs	*Increase number of workers*	*No other variables*	*Increase number of workers*	*No other variables*
Input Coefficient	0.29	0.17	0.136	0.399
Input Coefficient Interpretation	Every average increase of 9 days for which workers are hired, is predicted to have 135 451 shillings increase in average income in cluster	Every 3.5 average decrease in types of crops planted, is predicted to have 91 925 shillings increase in average income in cluster	Every 17 additional workers hired, on average, is predicted to have 270 470 shillings increase in average income in cluster	Every average decrease of 17 in types of crops planted, is predicted to have 1 369 790 shillings increase in average income in cluster
Other Impactful Input….	**Increase amount of pesticide used**	**Increase amount of pesticide used**	**Decrease crop diversification**	**Increase number of tools used**
…. which is highly correlated with these inputs	*Use of chemical fertiliser*	*Use of chemical fertiliser*	*No other variables*	*Use of chemical fertilizer*
Input Coefficient	0.262	0.153	0.125	0.206
Input Coefficient Interpretation	Every 1577 kg increase in pesticide used on average, is predicted to have 122 373 shillings increase in average income in cluster	Every 894 kg increase in pesticide used on average, is predicted to have 82 732 shillings increase in average income in cluster	Every 5.5 average decrease in types of crops planted, is predicted to have 248 594 shillings increase in average income in cluster	Every average increase of 4.5 in tools owned, is predicted to result in 707 210 shillings increase in average income in cluster

Cross-country comparisons

WORLD BANK GROUP

Cross-country comparisons

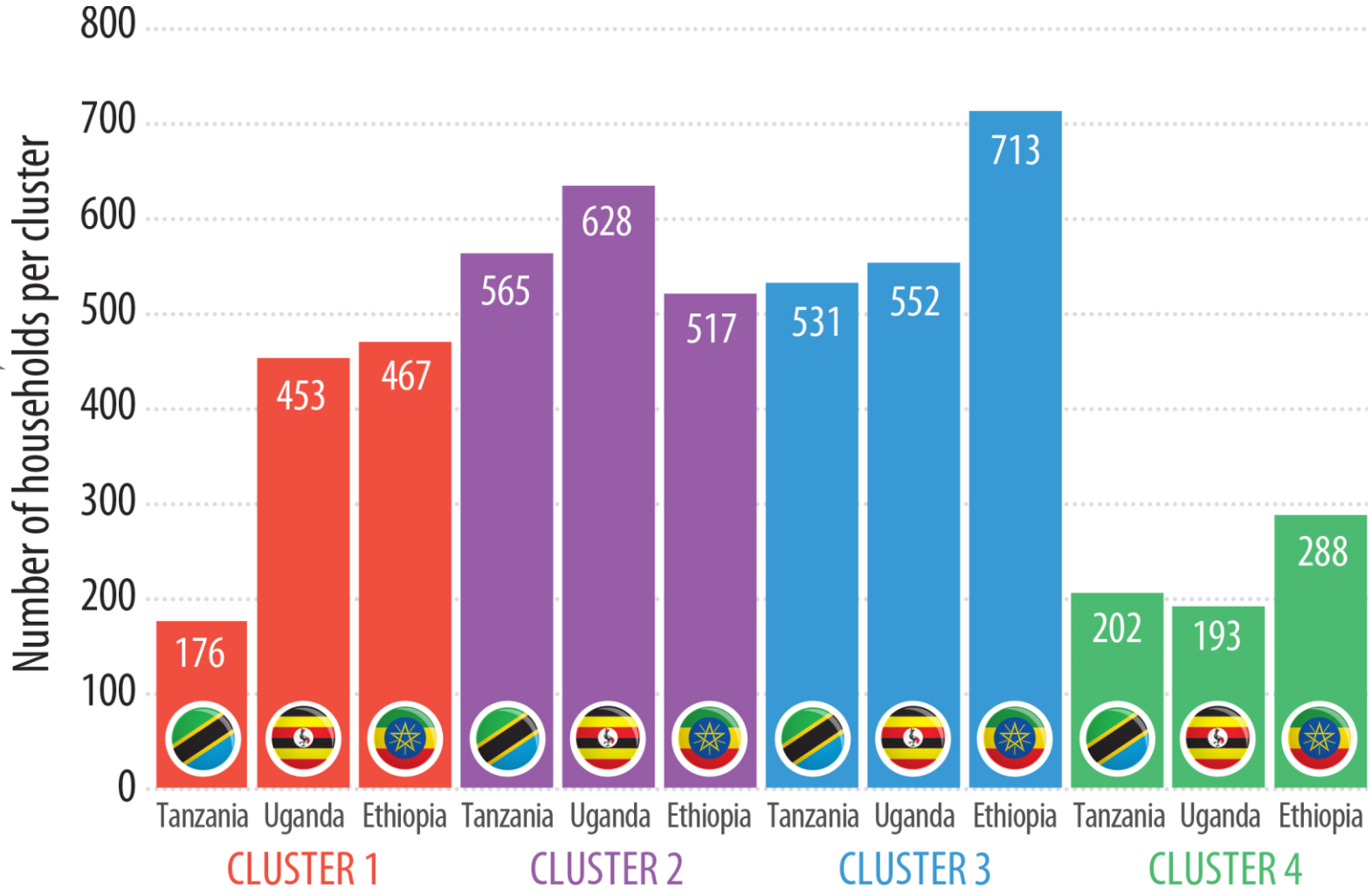
Cross-country comparison is limited by:

- Comparable data availability with low level of missingness across countries
- Difficulty to extract the same indicators across countries

EXAMPLE: Number of hired workers strong characteristic of highest cluster for Ethiopia

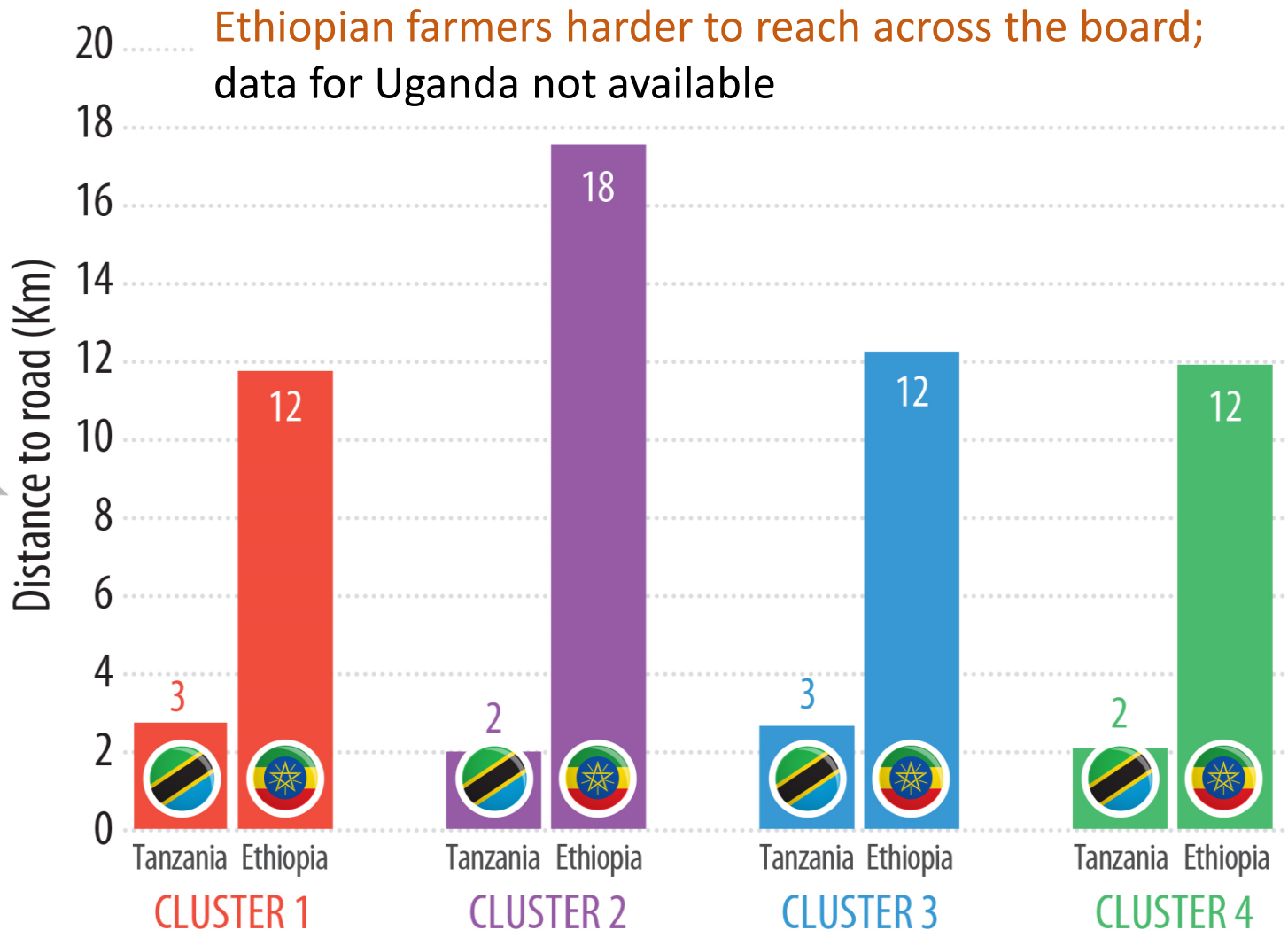
- Yet, in Tanzania's LSMS-ISA data set it is only measured as a binary (have you hired workers? yes/no, as opposed to number of workers hired)
- This does not allow us to measure the intensity margins observed with Ethiopia data

Country cluster comparison:
households per cluster



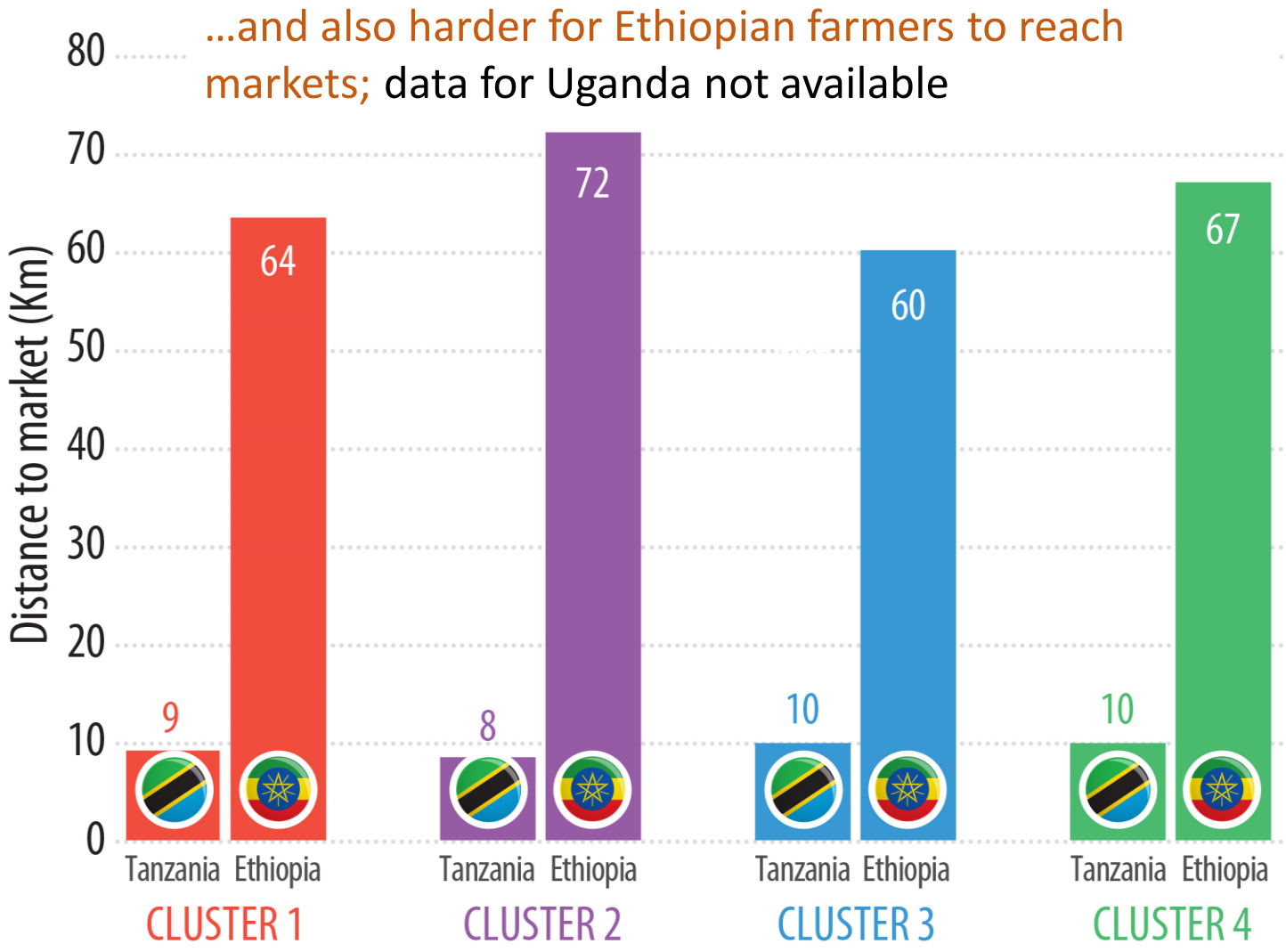
WORLD BANK GROUP

Country cluster comparison: distance to road (km)



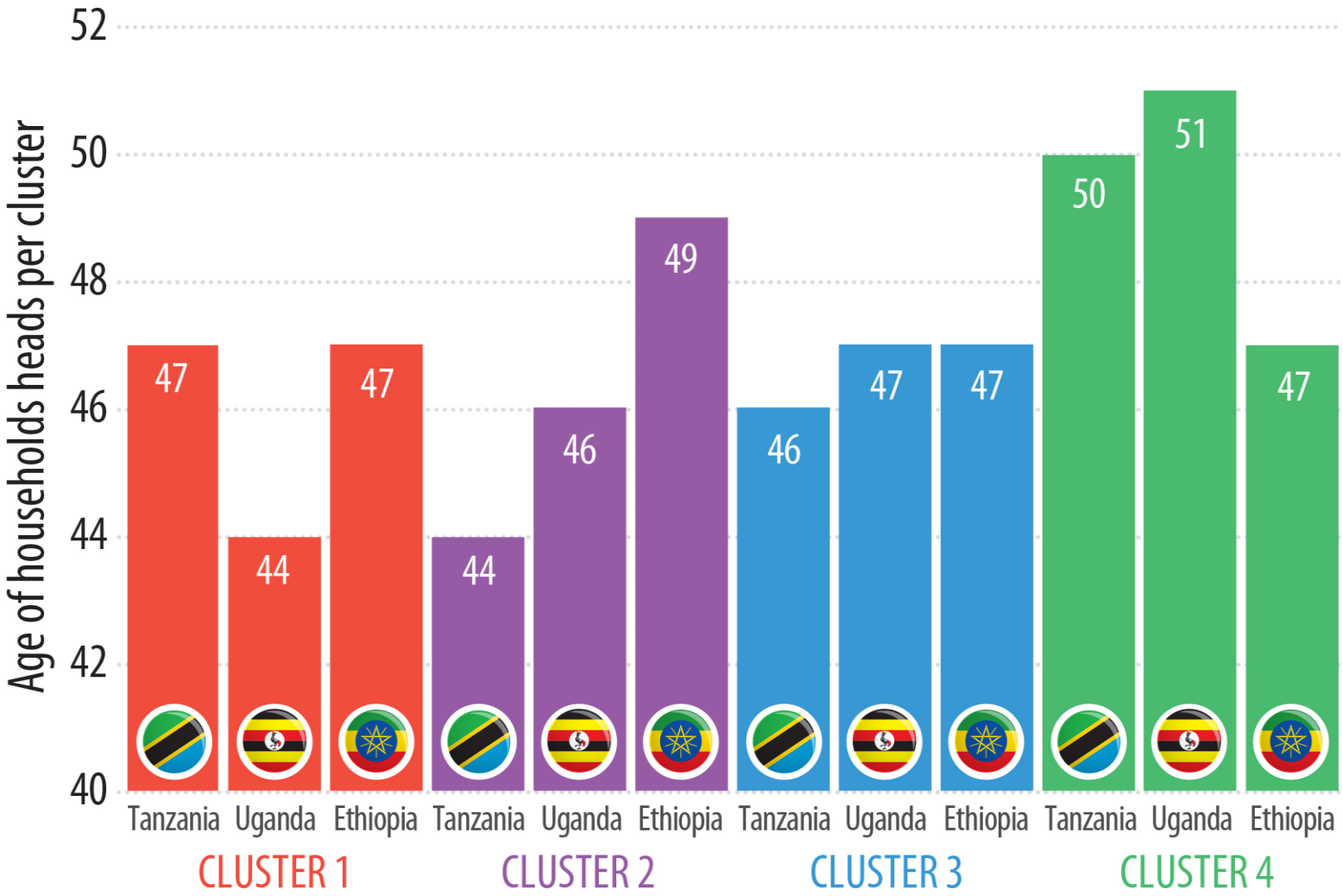
WORLD BANK GROUP

Country cluster comparison: distance to market (km)



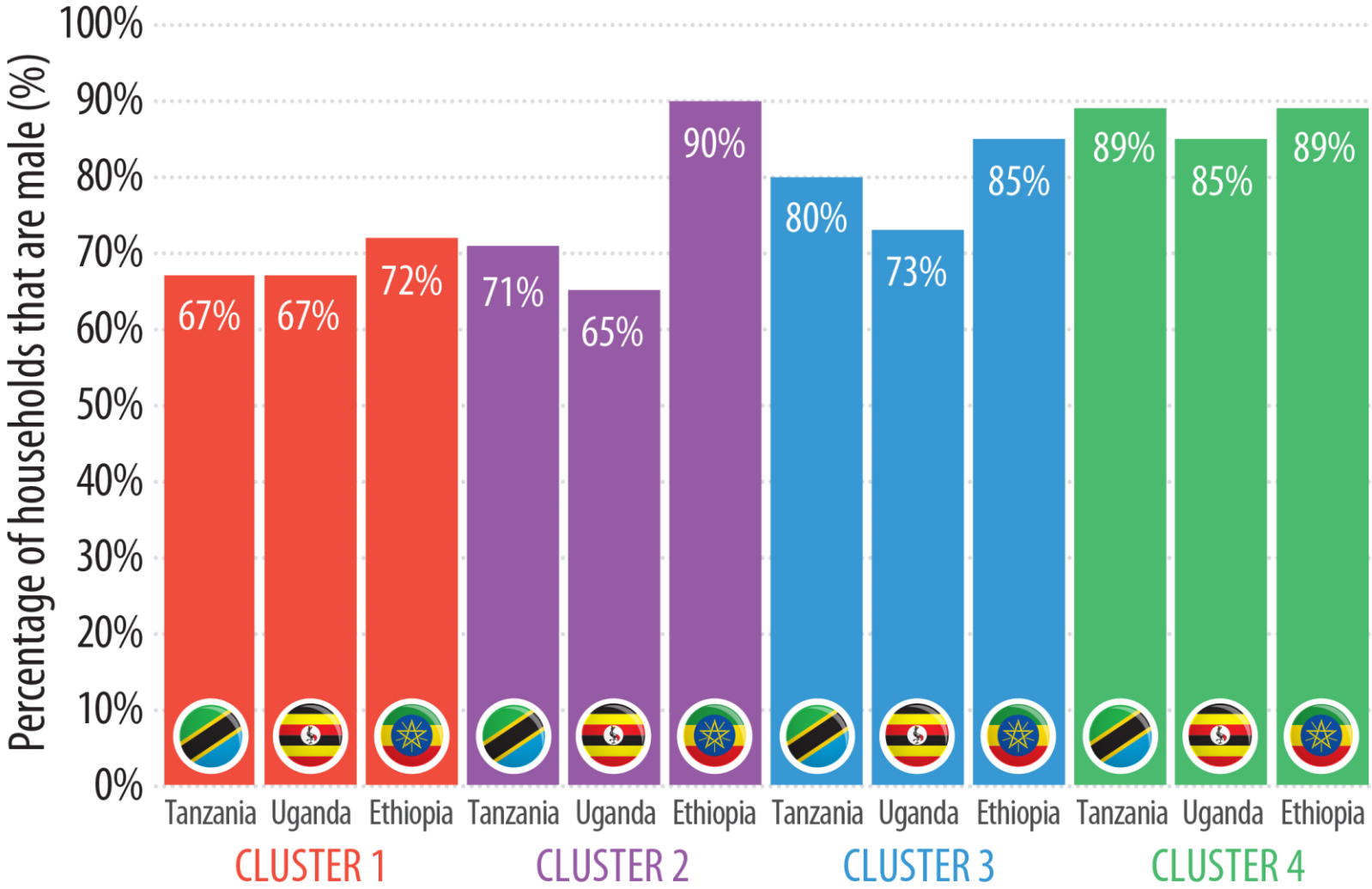
WORLD BANK GROUP

Country cluster comparison:
Age of household heads



WORLD BANK GROUP

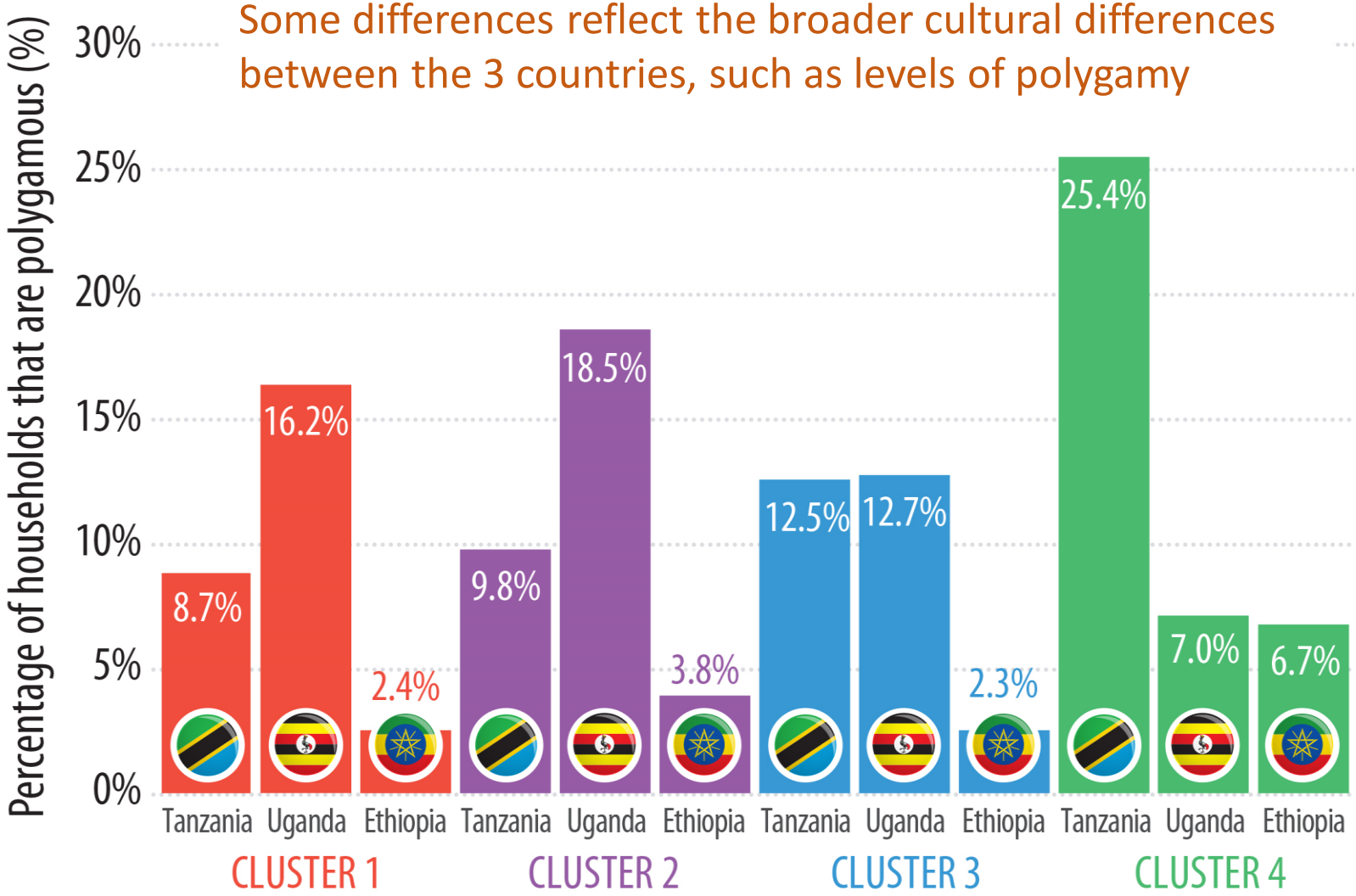
Country cluster comparison:
% households that are male



WORLD BANK GROUP

Country cluster comparison:
% households polygamous

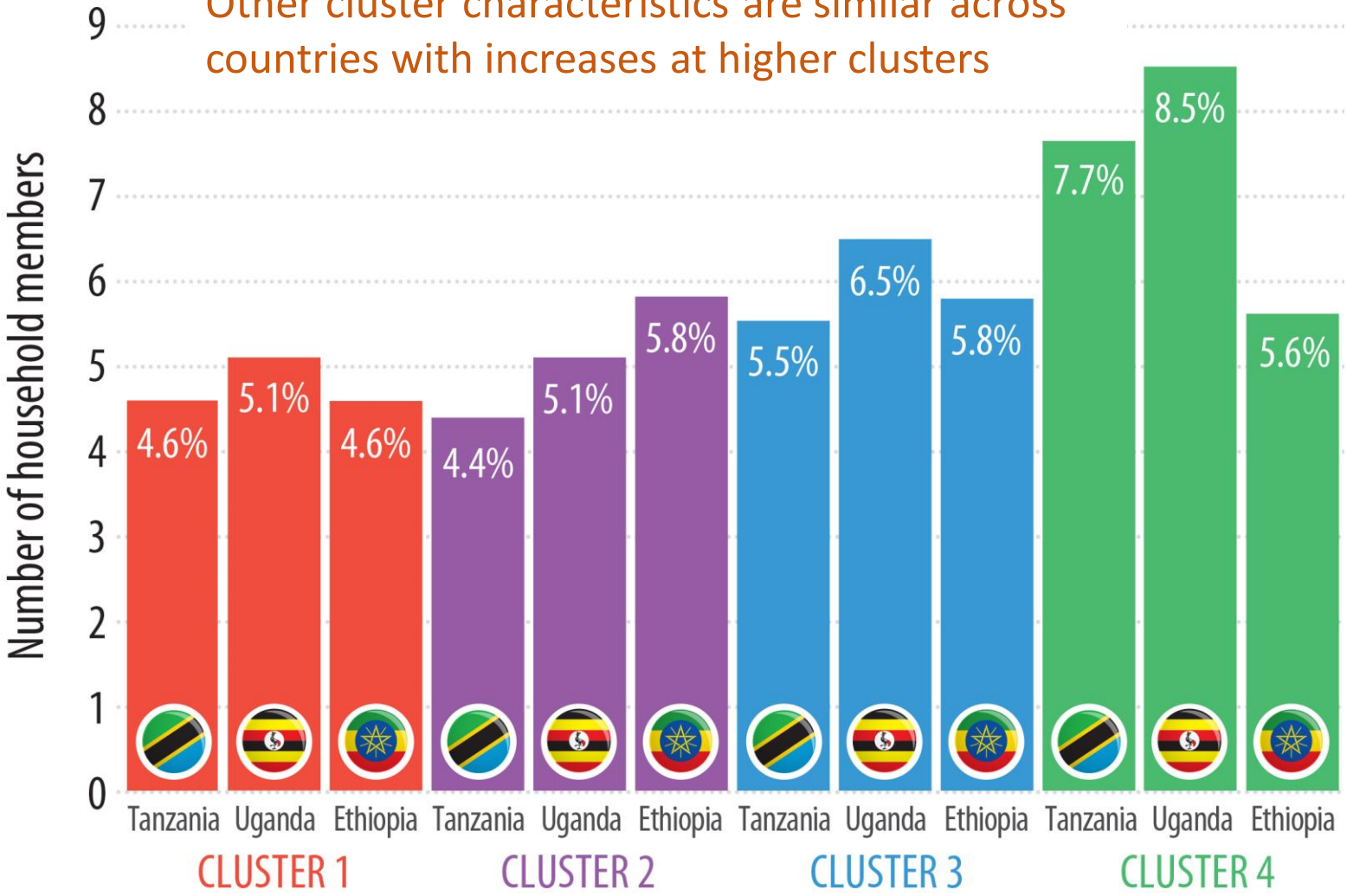
Some differences reflect the broader cultural differences between the 3 countries, such as levels of polygamy



WORLD BANK GROUP

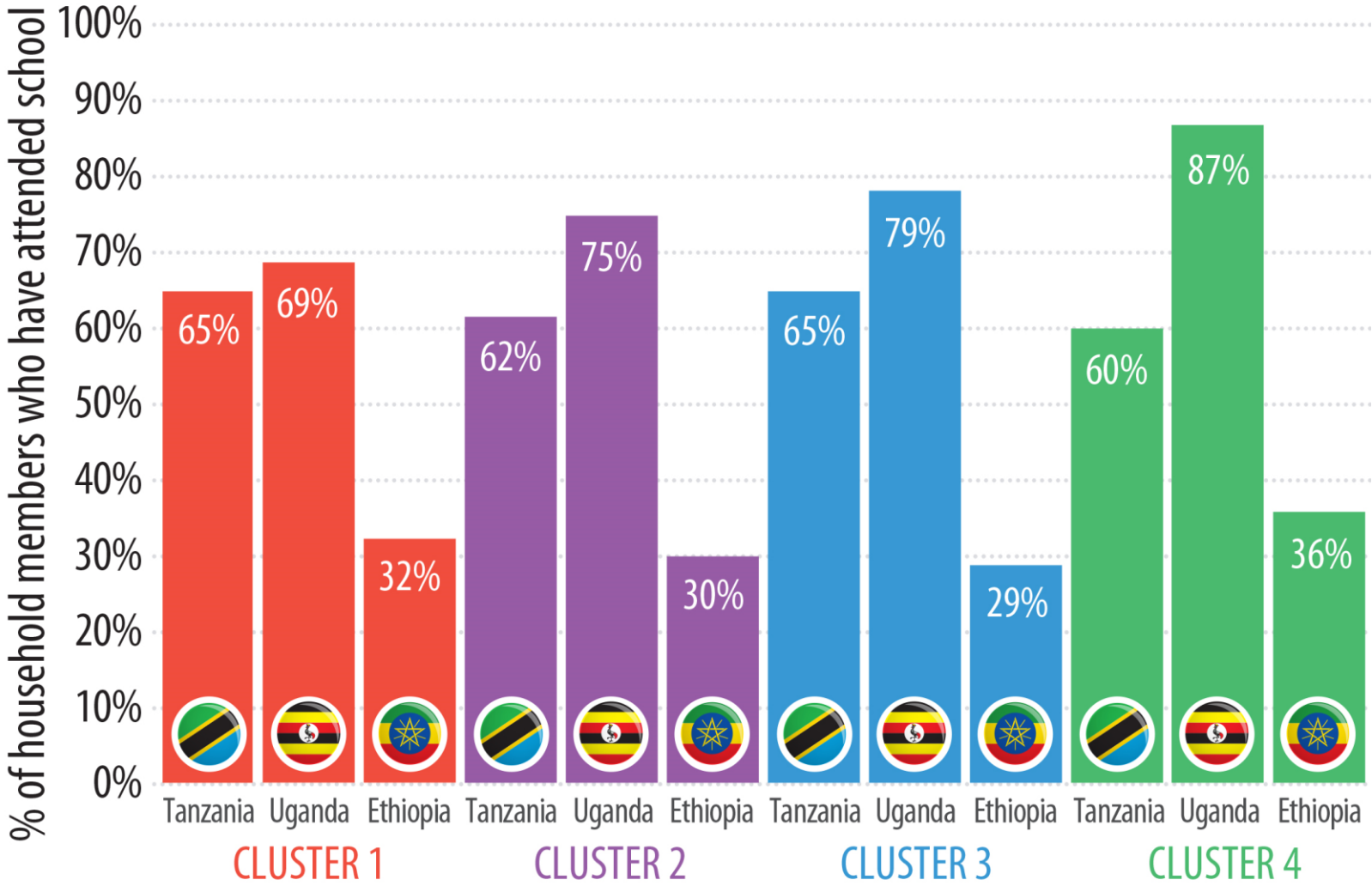
Country cluster comparison:
household members

Other cluster characteristics are similar across countries with increases at higher clusters



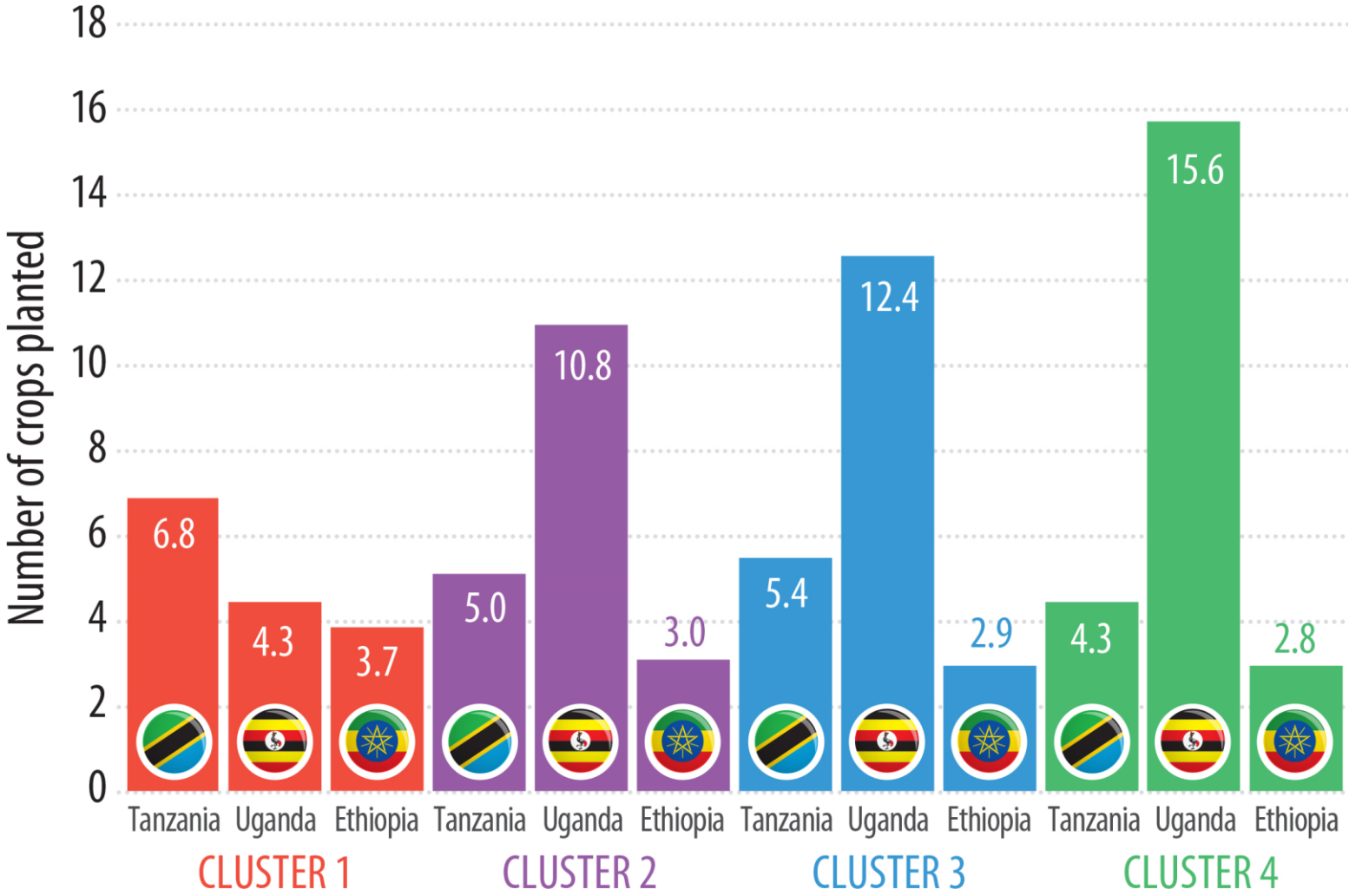
WORLD BANK GROUP

Country cluster comparison:
% household members who have attended school



WORLD BANK GROUP

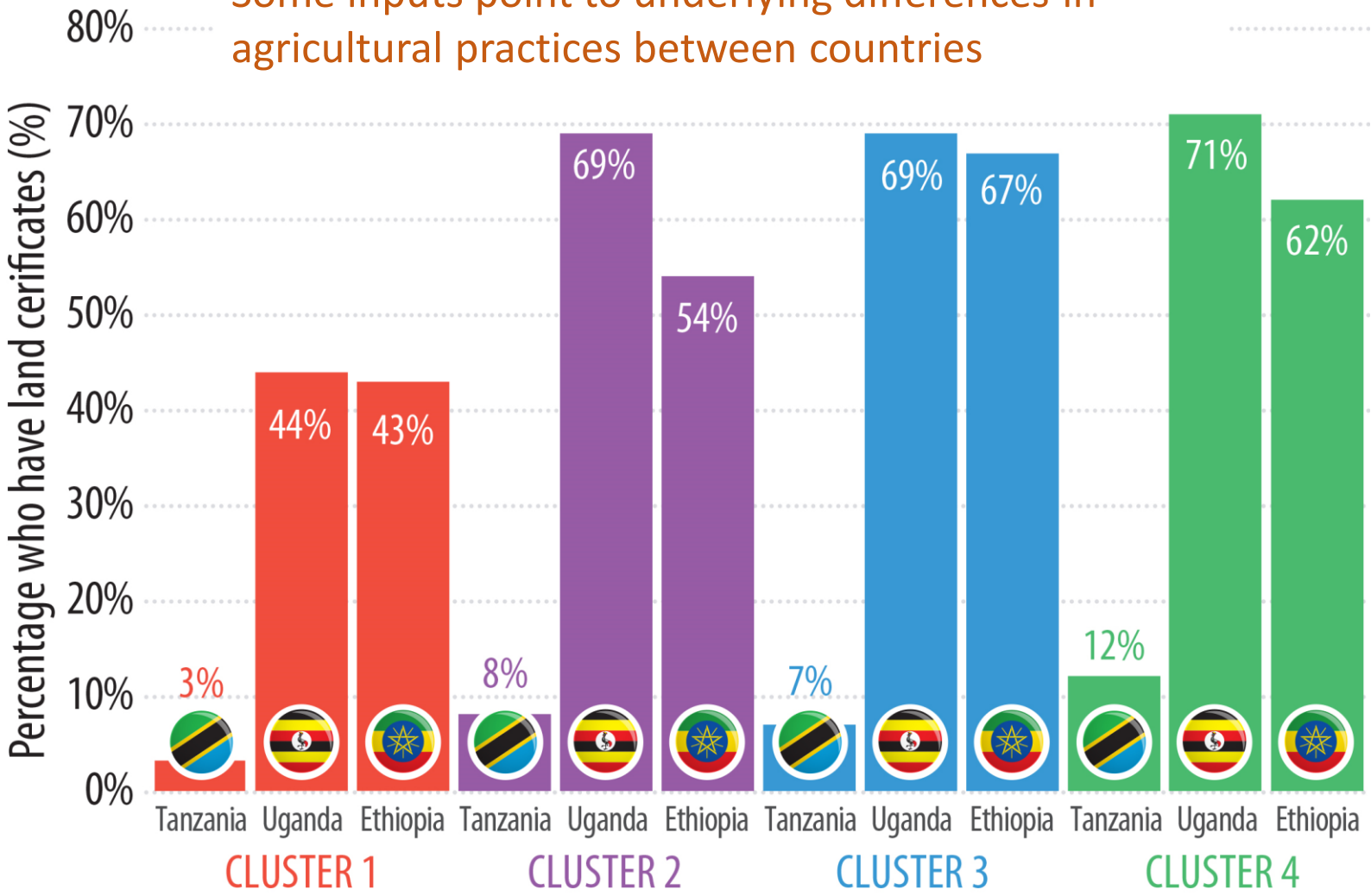
Country cluster comparison:
Number of crops planted



WORLD BANK GROUP

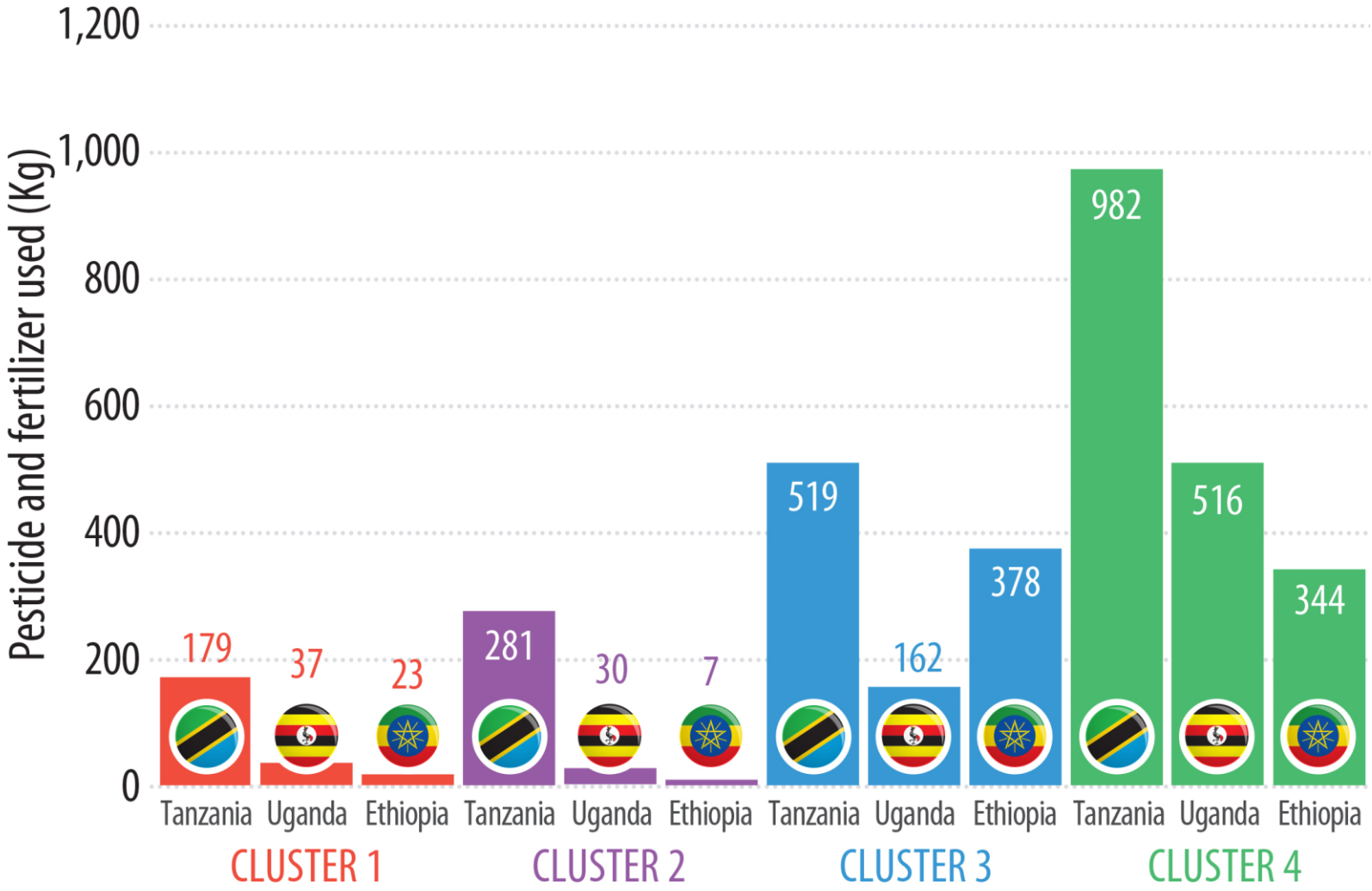
Country cluster comparison:
% who have land certificates

Some inputs point to underlying differences in agricultural practices between countries



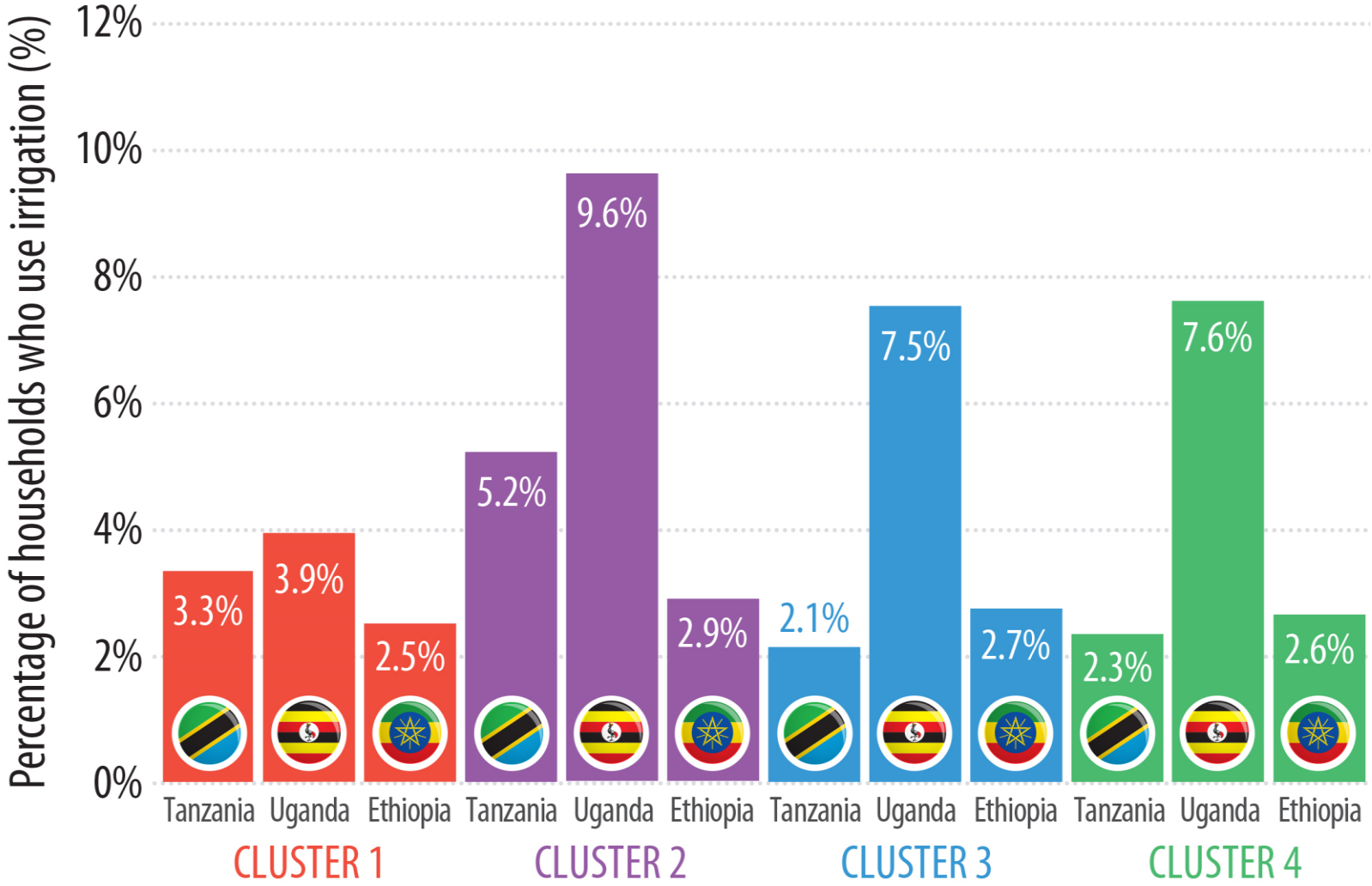
WORLD BANK GROUP

Country cluster comparison: pesticide and fertilizer use



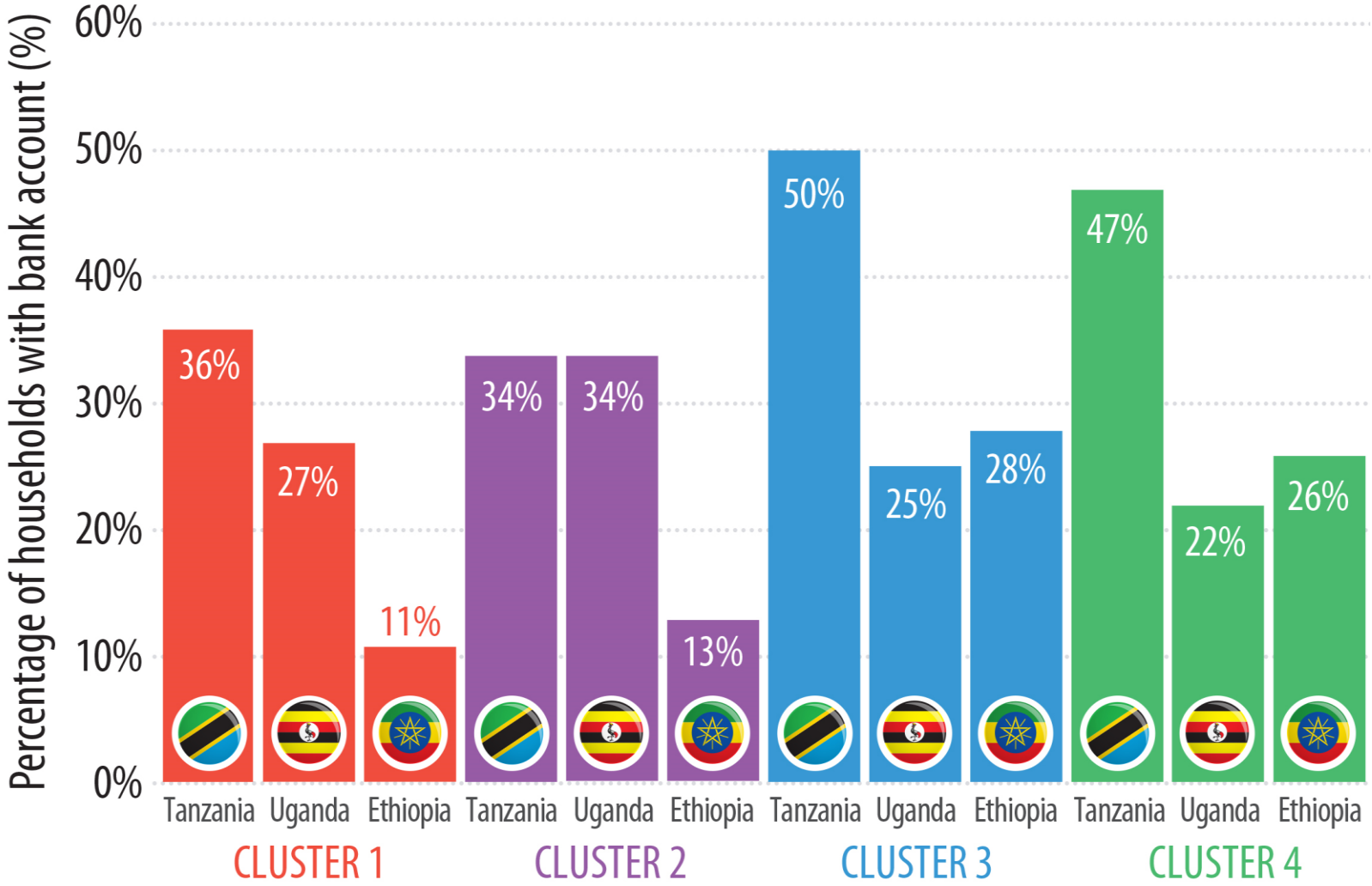
WORLD BANK GROUP

Country cluster comparison:
% of households who use irrigation



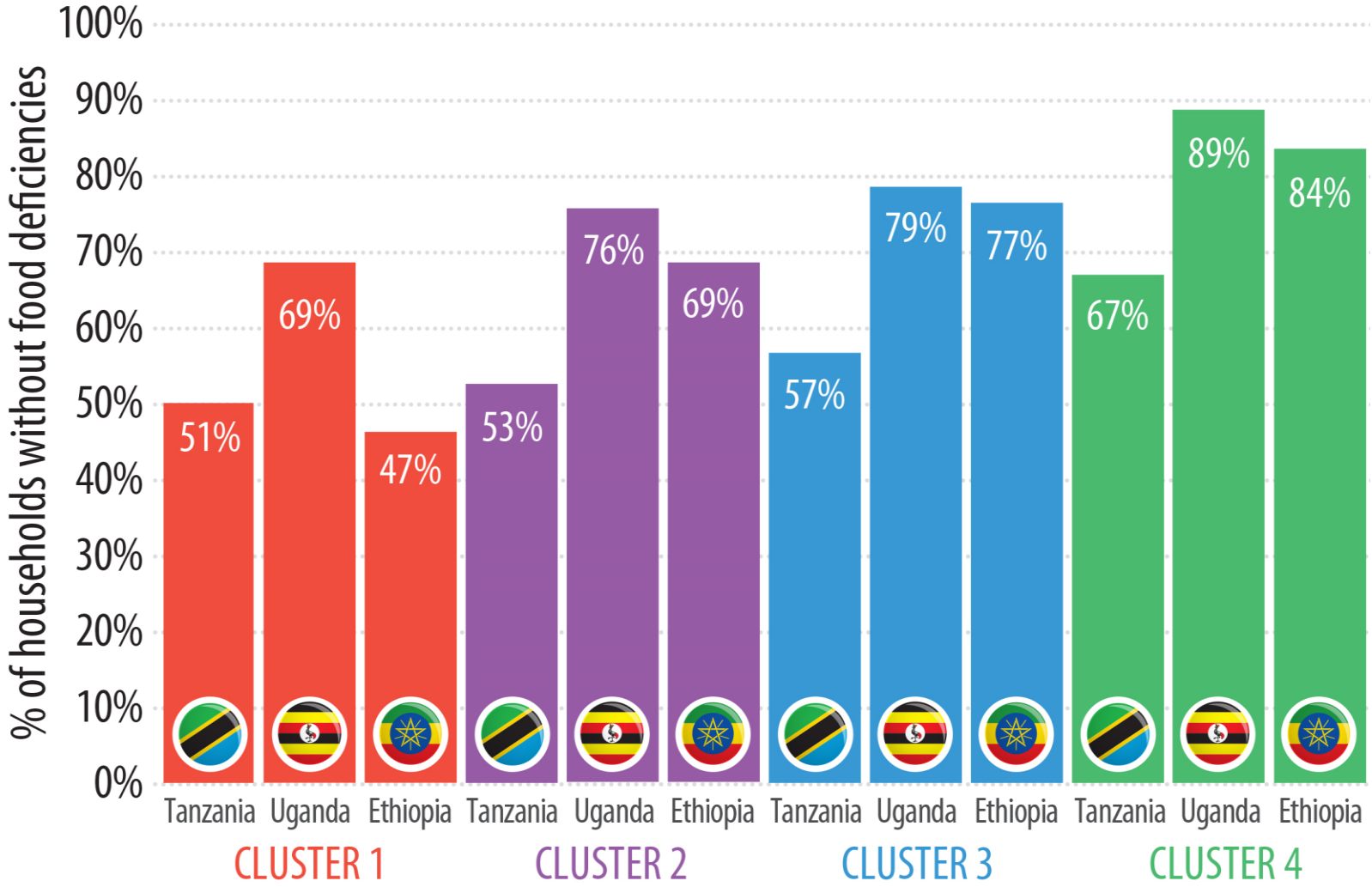
WORLD BANK GROUP

Country cluster comparison:
% with bank account



WORLD BANK GROUP

Country cluster comparison:
% without food deficiencies



WORLD BANK GROUP

Most impactful input: Comparison across countries

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
Most Impactful Input in Ethiopia	Increase farmers' savings	Increase # of hired workers	Increase # of hired workers	Increase # of hired workers
Most impactful input in Tanzania	Increase # of animals	Increase quantity of pesticide	Increase # of animals	Increase # of animals
Most impactful input in Uganda	Increase # of days for which workers are hired	Increase crop diversification	Increase number of days for which workers are hired	Increase crop diversification
Other Impactful Input in Ethiopia	Increase # of oxen owned	Obtain water storage pit	Increase quantity of chemical fertilizers used	Use extension program
Other impactful input in Tanzania	Increase quantity of pesticide	Decrease crop diversification	Increase quantity of pesticide	Increase quantity of pesticide
Most impactful input in Uganda	Increase quantity of pesticides used	Increase quantity of pesticides used	**Increase crop diversification**	Increase # of tools owned

Summary

WORLD BANK GROUP

Summary

We found a robust clustering of farmers in all 3 countries

- Characteristics associated with clustering in each country differ dramatically
- Clusters can be described as different phases of the agricultural transformation process
- Describes a pathway towards agricultural transformation
- Each inputs suggest a prioritized policy action at different phase of the transformation process

Summary

Most impactful input differs significantly between countries

Reasons include:

- Differences in correlations between inputs and outcomes
- Differences in farmer characteristics
- Differences in data
- Differences in underlying characteristics of population

Summary

Cross-country comparisons limited by lack of common measurement of some key inputs.

Yet, some patterns emerge:

- clustering analysis clearly shows that different farmers profiles exist across countries, suggesting to design cluster level policies
- inputs which are the most impactful of an increase in crop sales vary across clusters, supporting the implementation of cluster-level policies, rather than population level policies
- across countries, most predictive variables are hiring workers, usage of fertilizers or pesticides, animals, tools, irrigation, or animals; yet their relative importance across clusters (i.e., along income distribution) vary across countries
- interestingly the impact of crop diversification differs across country. Further analysis is required to show which specific crop leads to an increase in farmers competitiveness across countries