THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

LSMS
Living Standards Measurement Study

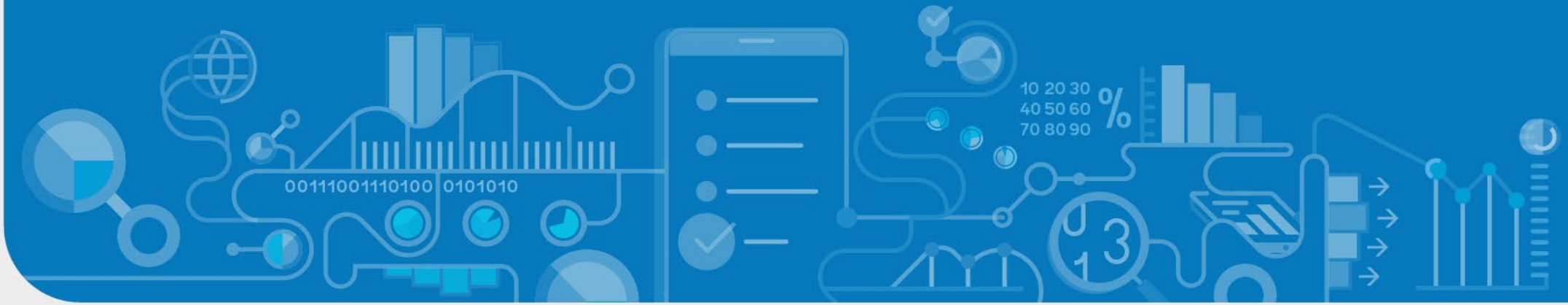
The World Bank's Flagship Household Survey Program

The Impact of Respondent Selection on Dietary Diversity and Quality Measurement in Surveys: Findings in Ethiopia and Nigeria

Alemayehu Ambel[1], Josh Brubaker[2], Talip Kilic[1] & Asmelash Tsegay[1]

1. Survey Unit, DECDG, World Bank
2. International Fund for Agricultural Development

Better Data for Better Jobs and Lives: Innovations in Survey Measurement in the Age of AI
8-9 December 2025



Motivation

- Dietary diversity and quality are important predictors of health and wellbeing (Kant, 2004; Martin-Prevel et al., 2017; Ruel, 2003; Verger et al., 2021).

- They are commonly assessed using indices, which are recognized as effective measures of wellbeing, especially in developing countries, due to their simplicity and ease of large-scale application (Herforth et al., 2020; Ruel, 2003; Trijsburg et al., 2019).

    - These indicators are constructed from individual-level data - recall of consumption of selected food items over a period of time- past 24 hours or yesterday (FAO & FHI 360, 2016; Herforth, Ballard & Rzepa, 2024; WDDP Study Group, 2017).

- One of these indicators, Minimum Dietary Diversity for women (MDD-W), was formally adopted as a new SDG (SDG 2.2.4) indicator by the UN Statistical Commission at its 56th session (March 2025) in New York.

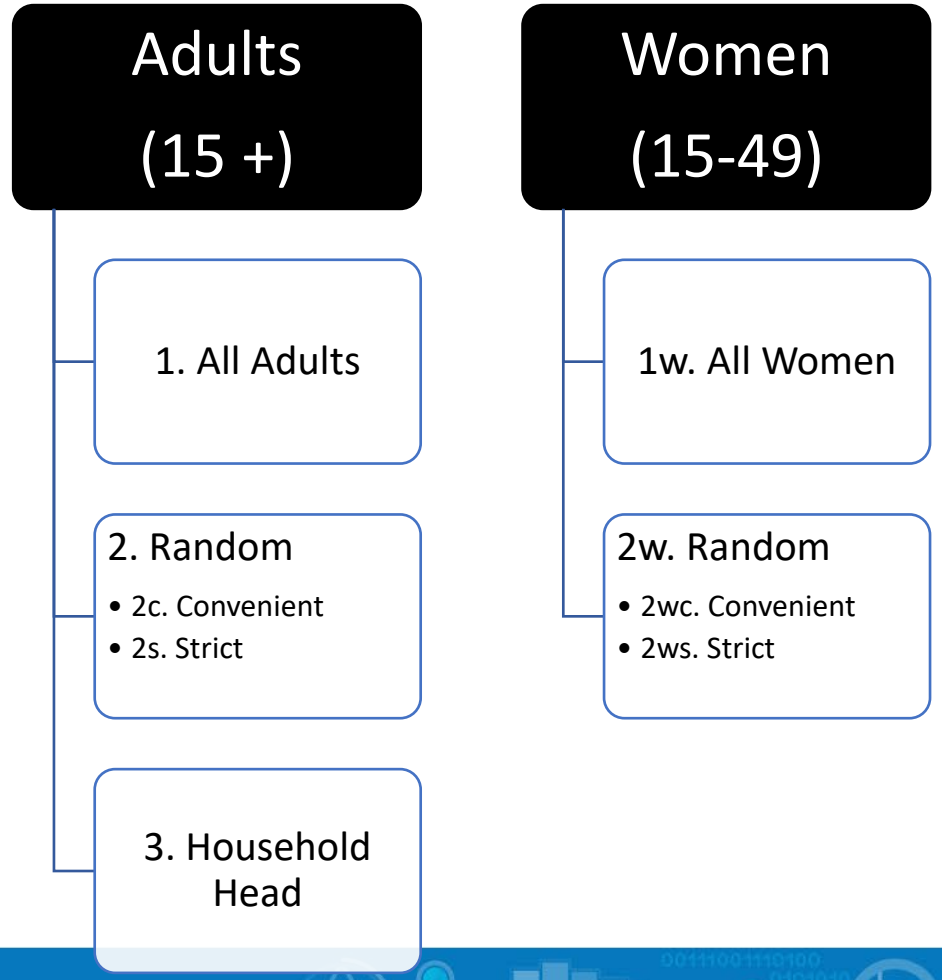
Research Questions

- Our research examines the data collection protocols for this indicator in household surveys, focusing on how eligible respondents are selected within households, using recent survey data from Ethiopia and Nigeria.

- Various protocols exist for collecting information from eligible individuals, each with its own benefits, limitations, and effects on the indicators.
  - For instance: interviewing every eligible member, randomly selecting one individual, interviewing the household head, or using a proxy respondent.

- Specifically, we address the following questions.
  - **Can only one eligible individual be selected at random to derive estimates that are equivalent of interviewing all eligible individuals?**
  - **Can the randomly selected individual be replaced in the event of unavailability?**
  - Are proxy respondents effective in filling data gaps when the intended interviewees cannot be reached?
  - **What are the implications for interviewing the household head ?**

Experimental Design: Sample Scenarios



Experimental Design: Re-weighting for sample reductions

Method 1: Simple correction for excluded observations to extrapolate estimates to the population of interest

1. Inclusion/exclusion of part of the pre-adjustment sample.

$$P \begin{cases} I|Rule = 1 \\ J|Rule = 0 \end{cases}$$

- where *P* is the survey population pre-adjustment, which is divided by a rule into sets *I* where rule=1 and *J* where rule=0.
- Rule can include the set of respondent selection scenarios, and also the protocol followed for proxy response

2. Modify weights for an individual in population *I* as follows.

$$w_{i,a} = (w_{i,p}|\text{Rule} = 1) * \left[1 + \left(\frac{\sum_{j=1}^{J} w_{j,p}|Rule=0}{\sum_{p=1}^{P} w_p}\right)\right]$$

- where $w_{i,a,}$ is an adjusted weight and $w_{i,p}$ is the unadjusted weight for observation *i*.

3. Post-stratify according to the sum of weights in the baseline survey.

Experimental Design: Re-weighting for Scenarios

Method 2: Estimate individual probability to be available to respond for themselves. Use this probability ($\widehat{self}$) to weight responses more highly that had lower probability to be available to respond.

1. Estimate a logistic regression

$$\Pr(self = 1) = F(\beta_0 + \sum_{k=1}^{K} \beta_k X_k)$$

- where the dependent variable is a binary that is equal to one where a household member was self-respondent and 0 otherwise;
- X is a vector containing K independent explanatory variables including individual characteristics, household characteristics and a spatial fixed effect.

2. Create adjustment factor using the estimated probability to be a self respondent ($\widehat{self}$),

3. Construct the deciles of estimated probability and calculate the average probability within decile then take the inverse of that average .

$$af_{D=d} = \frac{1}{\frac{\sum_{i=1}^{N} \widehat{self}_i}{N}}$$

Experimental Design: Overall framework

Scenario	Sample	Adults		
	Self-report protocol	**Proxy included**	Require self-report	
	Adjustment protocol		**Unadjusted**	**Adjusted**
1.	All Adults	Full sample (Adults)	Self=1 * Method 1	Self=1 * Method 1 & 2
2c.	Convenience Random	Random=1 * Method 1	Self=1, then Random=1 * Method 1	Self=1, then Random=1 * Method 1 & 2
2s.	Strict Random	Random=1 * Method 1	Random=1, then Self=1 * Method 1	Random=1, then Self=1 * Method 1 & 2
3.	Household Head	Head=1 * Method 1	Head=1 & Self=1 * Method 1	Head=1 & Self=1 * Method 1 & 2

	Women (15-49)		
	Proxy included	Require self-report	
		Unadjusted	**Adjusted**
1w.	Full sample (Women)	Self=1 * Method 1	Self=1 * Method 1 & 2
2wc.	Random=1 * Method 1	Self=1, then Random=1 * Method 1	Self=1, then Random=1 * Method 1 & 2
2ws.	Random=1 * Method 1	Random=1, then Self=1 * Method 1	Random=1, then Self=1 * Method 1 & 2

Experimental Design: Respondents

Country	Target Population	Sample	Selection Scenario	Proxy Included	Unadjusted Self-Report	Adjusted Self-Report
Ethiopia	Adults (15+)	**1**	**All Adults**	14,058	8,802	**8,665**
		2c	Convenience Random		4,914	4,792
		2s	Strict Random		3,431	3,313
		3	Household head	4,950	3,653	3,534
	Women (15-49)	**1W**	**All Women**	6,218	4,352	**4,352**
		2Wc	Convenience Random		3,602	3,602
		3Ws	Strict Random		3,214	3,214
Nigeria	Women (15-49)	**1W**	**All Women**		4,330	**4,330**
		2Wc	Convenience Random		3,032	3,032
		3Ws	Strict Random		2,710	2,710

Results: Comparison between self-report and proxy (Adults 15+ sample) – Individual respondent characteristics

Variable (Individual characteristics)	Self-Report =1	Self-Report =0	Difference	
Age†	36.54	30.18	-17.4	***
Age 15-24	0.23	0.49	25.9	***
Age 25-49	0.57	0.39	-18.2	***
Female	0.57	0.4	-16.4	***
Head	0.4	0.23	-17.1	***
Spouse of head	0.32	0.11	-20.8	***
Child of head	0.21	0.52	30.5	***
Married	0.65	0.37	-27.3	***
No education	0.44	0.31	-12.2	***
Primary education	0.36	0.45	8.9	***
Secondary education	0.12	0.17	5.1	***
Tertiary education	0.08	0.06	-1.9	**
Can read and write	0.55	0.68	12.9	***
Owns working cellphone	0.27	0.24	-3.3	*
Worked in agriculture	0.45	0.45	0	
Worked in non-farm enterprise	0.08	0.07	-1.2	
Worked for wages	0.15	0.15	0.3	

Results: Comparison between self-report and proxy (Adults 15+ sample) – Household characteristics

Variable (Housheold Characersitics)	Self-Report =1	Self-Report =0	Difference	
Any income-generating work	0.6	0.6	-0.5	
Dependency ratio†	0.74	0.59	-20.6	***
Eligible adults†	3.46	4.25	22.8	***
Eligible females†	1.42	1.75	23.3	***
Lowest wealth quintile	0.3	0.29	-0.5	
Second wealth quintile	0.31	0.33	1.5	
Third wealth quintile	0.23	0.19	-3.6	*
Fourth wealth quintile	0.1	0.11	1.1	
Highest wealth quintile	0.06	0.08	1.6	
No land ownership	0.22	0.23	0.7	
Smallest quintile of land area	0.08	0.06	-1.7	
Second quintile of land area	0.11	0.1	-1.5	
Third quintile of land area	0.17	0.17	0.1	
Fourth quintile of land area	0.19	0.15	-3.8	**
Largest quintile of land area	0.23	0.29	6.3	**

Results: Comparison between self-report and proxy (Women sample) – Individual characteristics

	Ethiopia				Nigeria			
Variable (Individual characteristics)	Self-report=1	Self-report=0	Difference		Self-report=1	Self-report=0	Difference	
Age†	30.8	24.36	-20.9	***	31.74	24.56	-22.6	***
Age 15-24	0.28	0.63	35.7	***	0.3	0.63	33.8	***
Age 25-49	0.72	0.37	-35.7	***	0.7	0.37	-33.8	***
Female	1.00	1.00	0.00		1.00	1.00	0.00	
Head	0.13	0.02	-10.5	***	0.1	0.01	-6.3	***
Spouse of head	0.61	0.23	-38.2	***	0.6	0.33	-27.6	***
Child of head	0.18	0.58	39.7	***	0.2	0.52	27.4	***
Married	0.72	0.31	-41	***	0.6	0.35	-28.6	***
No education	0.45	0.31	-13.8	***	0.3	0.26	-6.8	**
Primary education	0.37	0.46	8.9	***	0.3	0.3	2.3	
Secondary education	0.11	0.18	6.7	***	0.3	0.31	4.3	*
Tertiary education	0.07	0.05	-1.8	*	0.1	0.13	0.2	
Can read and write	0.53	0.67	14.1	***	0.7	0.76	9.3	***
Owns working cellphone	0.19	0.14	-5.1	***	0.6	0.55	-2.9	
Worked in agriculture	0.33	0.35	2		0	0.02	-1.5	**
Worked in non-farm enterprise	0.08	0.07	-1.3		0.3	0.18	-15.1	***
Worked for wages	0.11	0.1	-1.4		0.1	0.06	-2.4	**

Results: Comparison between self-report and proxy (Women sample) –Household characteristics

	Ethiopia				Nigeria			
Variable (Housheold Charactersitics)	Self-report=1	Self-report=0	Difference		Self-report=1	Self-report=0	Difference	
Any income-generating work	0.49	0.48	-0.9		0.4	0.25	-18.9	***
Dependency ratio†	0.81	0.52	-35.7	***	1	0.75	-22.8	***
Eligible adults†	3.34	4.48	34.2	***	3.4	4.27	23.8	***
Eligible females†	1.67	2.34	40.7	***	2	2.64	30.7	***
Lowest wealth quintile	0.28	0.32	3.7		0.2	0.13	-4.5	**
Second wealth quintile	0.31	0.32	0.6		0.2	0.16	-6.2	***
Third wealth quintile	0.23	0.17	-5.8	**	0.2	0.19	-0.3	
Fourth wealth quintile	0.12	0.11	-0.4		0.2	0.24	4.8	**
Highest wealth quintile	0.07	0.08	1.8		0.2	0.28	6.2	***
No land ownership	0.23	0.24	1		0.2	0.16	0.6	
Smallest quintile of land area	0.08	0.06	-1.5		0.2	0.16	0.2	
Second quintile of land area	0.11	0.09	-1.8		0.1	0.13	-0.6	
Third quintile of land area	0.17	0.18	0.9		0.2	0.19	1.3	
Fourth quintile of land area	0.19	0.16	-3		0.2	0.16	-1.6	
Largest quintile of land area	0.22	0.27	4.5		0.2	0.19	0.1	

Results: Imbalanced covariates & DDQ outcomes

	Ethiopia						Nigeria			
	Adults		Women (15-49)				Women (15-49)			
	Dietary Diversity Score		Food groups consumed by women		MDD-W		Food groups consumed by women		MDD-W	
Age	-0.021	*	-0.11	**	0		-0.03		0	
Age squared	0	*	0.002	**			0			
Sex = Female	-0.121									
Relation to Head = Spouse	0.162		0.297	*	0.072	**	0.235		0.01	
Relation to Head = Child	-0.291	*	-0.14		-0.01		0.246		0.044	
Relation to Head = Other	0.003		0.443		0.104	*	0.312		0.053	
Education = Primary	0.334	**	0.148		-0.01		-0.25	*	-0.08	**
Education = Secondary	0.661	***	0.408	*	0.038		-0.05		-0.01	
Education = Tertiary	0.702	***	0.313		0.019		0.146		0.024	
Marital status = Married	-0.088		0		-0.03		0.106		0.066	
Income generating work	0.153		0.26	*	0.027		0.445	***	0.077	***
Eligible adults	0.075									
Eligible females			-0.03		0		0.044		0.032	***
Dependency ratio	0.01		-0.05		-0.01		-0.02		0	
Household member can read and write in any language	-0.079		0.132		0.082	**	0.194		0.061	*
Constant	2.744	***	3.933	***			4.107	***		
Spatial Fixed Effects	Yes		Yes		Yes		Yes		Yes	
Number of Observations	8666		4352		4352		4330		4330	
Model	OLS		OLS		Logit		OLS		Logit	
Overall F-test	5.65		5.01		3.55		7.95		5.73	
Prob > F	0		0		0		0		0	

Results: Likelihood of being self-respondent

	Ethiopia				Nigeria	
	All adults (15+)		Women		Women	
Age = 15-24	-0.08	***				
Age = 50+	0.004					
Age			0.006	***		
Age (ln)					0.217	***
Sex = Female	0.11	***				
Relation to Head = Spouse	0.033		-0.138	***	-0.158	***
Relation to Head = Child	-0.256	***	-0.4	***	-0.131	***
Relation to Head = Other	-0.258	***	-0.368	***	-0.174	***
Marital status = Married			0.101	***	0.061	
Education = Primary			0.079	***	0.063	**
Education = Secondary			0.048		0.033	
Education = Tertiary			0.11	***	0.039	
Worked in agriculture			-0.056	*	0.122	**
Worked in non-farm enterprise			-0.068	*	0.05	**
Worked for a wage			-0.01		0.022	
Eligible adults	-0.023	***				
Eligible females			-0.032	***	-0.046	***
Dependency ratio	-0.021					
Wealth quintile = Second	-0.003				-0.008	
Wealth quintile = Third	-0.004				-0.066	**
Wealth quintile = Fourth	-0.091	*			-0.091	***
Wealth quintile = Highest	-0.084				-0.095	***
Land ownership = Lowest quintile	0.128	**	0.158	**		
Land ownership = Second quintile	0.056		0.151	***		
Land ownership = Third quintile	0.04		0.131	**		
Land ownership = Fourth quintile	0.1	*	0.159	***		
Land ownership = Highest quintile	0.04		0.116	*		
Spatial Fixed Effects	Region x Urban		Region x Urban		Zone x Urban	
Number of Observations	13908		6218		5490	
Overall F-test	10.29		9.38		11.08	
Prob > F	0		0		0	

Results: Balance summary on covariates by survey response protocol & sample scenario

Self-Response Protocol:	Proxy Included			Unadjusted Self Report				Adj Self Report		
Scenario for Adult (15+)	1	2	3	1	2c	2s	3	2c	2s	3
Ethiopia	93	101	371	91.7	82.4	94	343	19	54.4	339
Scenario for Women(15-49)	1W	2W		1W	2Wc	2Ws		2Wc	2Ws	
Ethiopia	33.6	41.8		87.6	85	108		6.9	43.4	
Nigeria				61.1	47.9	73.6		5.1	32	

Results: Impact on Dietary Diversity and Quality Outcomes

			Benchmark	Alternatives													
		Self-Response Protocol:	Adj Self Report	Proxy Included					Unadjusted Self-Report						Adjusted Self-Report		
Country	Sample	Scenario	1 (All Adults)	1		2		3		1		2c		2s		3	2c		2s		3
Ethiopia	Adults (15+)	All-5: Consumed all five recommended food groups	0.052	-0.3		-1.0	*	-0.8		-0.3		-0.6		-1.0	*	-0.8	-0.7	*	-0.8		-0.4
		Dietary Diversity Score (DDS)†	3.011	-2.0	***	-3.2	***	-2.3	*	-1.0		-1.1		-2.7	**	-0.7	-0.9		-2.0		0.4
		NCD-Protect total score†	2.539	-2.5	***	-3.8	***	-4.0	***	-1.4	*	-1.7	*	-3.0	**	-2.7	-1.2		-2.1		-1.6
		NCD-Risk total score†	0.369	-5.9	**	-1.7		-7.1	*	-6.0	***	-1.5		-3.6		-2.1	2.5		2.5		3.6
		GDR Score†	11.17	-0.4	*	-0.8	***	-0.7	*	-0.1		-0.3		-0.6		-0.6	-0.4		-0.6		-0.5
		Scenario	1W (All Women)	1W		2W				1W		2Wc		2Ws			2Wc		2Ws		
	Women (15-49)	Food groups (0-10) consumed yesterday by women aged 15-49†	2.987	-1.0		0.3				-0.8		-0.7		-1.2			0.3		0.6		
		Minimum dietary diversity for women (5+ food groups)	0.166	-0.5		0.2				-0.9	•	-0.9		-0.1			0.0		1.0		
Nigeria	Women (15-49)	Food groups (0-10) consumed yesterday by women aged 15-49†	4.324							-0.4		0.0		0.8			0.6		1.3	•	
		Minimum dietary diversity for women (5+ food groups)	0.432							-0.5		-0.4		-0.4			0.0		-0.2		

Summary of Results

- **Can only one eligible individual be selected at random to derive estimates that are equivalent of interviewing all eligible individuals?**

- Yes, both approaches of random selection provided better results in terms of mimicking the benchmark of interviewing all.

- **Can the randomly selected individual be replaced in the event of unavailability?**

- Yes, in our case, provided the adjustment considers the likelihood hood of being self-respondent.

Summary of Results

- **Can proxy respondents accurately fill data gaps when interview targets cannot be reached?**
- No, they are different, shouldn't be included and the weights need to be adjusted accordingly for non-response.
- **What are the implications for interviewing the household head ?**
- Not consistent because it provided the largest differences in terms of imbalanced covariates. But not significant difference on dietary quality outcomes with PS adjustment.

Limitations

- **This study is based on a post-fieldwork design.**
  - The respondent selection choices are based on what could be done with the data from these two surveys.
  - The proxy respondent analysis is indirect using common observables, no self and proxy response was collected on the same respondent contemporaneously.
- **The analysis is context specific.**
  - Focus is specifically on dietary quality recall and may be relevant to similar subjects/contexts.
  - However, it should not be generalized to other subjects. Proxy responses might be suitable for different topics

LSMS
Living Standards Measurement Study

The LSMS website