



Food and Agriculture
Organization of the
United Nations

ADB

Quantifying the effects of CAPI versus PAPI using Sri Lanka's Agricultural Household Survey

Lakshman Nagraj Rao

Statistician

Statistics and Data Innovation Unit

Asian Development Bank

NagrajRao@adb.org

<https://www.adb.org/data/statistics>

Outline

- Introduction
- Research Objective
- What is a randomized experiment?
- Types of Errors and Definition of Total Error
- Empirical Strategy
- Results
- Conclusion
- Recommendation

Introduction

- Datasets matter in for policy analysis, yet underlying data collection process rarely considered
- Most of what we know about switching from PAPI to CAPI (from a research perspective) comes from experiences in Western countries.
- Literature from developed economies shows that:
 - Interviewers and interviewees react favourably to CAPI;
 - No problem for older respondents;
 - CAPI reduces errors; and
 - CAPI does not affect response rates

Introduction

- Not a great deal of evidence for
 - (i) the types of complicated agricultural surveys NSO's or other line ministries implement (**think 50x2030!**),
 - (ii) developing countries in Asia and the Pacific
- Fafchamps et al. (2010, JDE) show CAPI has no impact on precision of profits and sales measurement in Ghana.
- Caeyers et al. (2011, JDE) show CAPI impacts data quality and analysis for a household consumption survey in Tanzania.

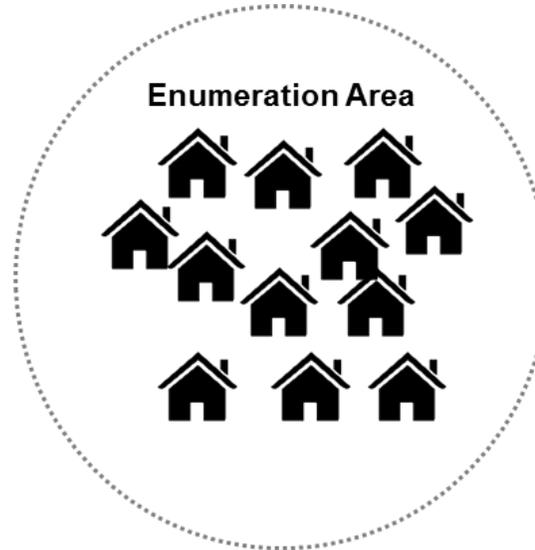
Research Objective

Using a **randomized** roll out of CAPI and PAPI Agricultural Household Survey (Maha season 2017) in Anuradhapura District, conducted as part of [ADB R-CDTA 9018](#) in partnership with the Department of Census and Statistics (Sri Lanka), the study aims to provide answers to the following questions:

1. Does CAPI help reduce the **number of errors** made in the questionnaire?
2. What are the **cost implications** of switching to CAPI?
3. How do respondents **perceive** CAPI?
4. What is the impact of CAPI on key agricultural statistics such as **rice yield**?

What is a randomized experiment?

-  Control (PAPI)
-  Treatment (CAPI)



Within each enumeration area, a random sample of households are allocated to the treatment and control group. Both groups are given the same survey questionnaire.



Households selected as treatment are interviewed using the **computer-assisted personal interviewing** method.



Households selected as control are interviewed using the **pen-and-paper interview** method.

Table 1. Sampling Characteristics

DESCRIPTION	Anuradhapura District, Sri Lanka	
	CAPI	PAPI
Total number of enumeration area	191	191
Total number of households interviewed	1,783	1,825
Average number of households per enumeration area	9.3	9.6
Total count of household members Interviewed	6,840	7,004
Average household size	3.8	3.9

Q1. Does CAPI help reduce the number of errors made in the questionnaire?

Definition of Total Survey Error

The sum of all type of errors committed by the household

$$\text{Total Survey Errors} = \text{Skip} + \text{Data Validation} + \text{Missing} + \text{Logic}$$

1. Skip Error

ONLY FOR HOUSEHOLD MEMBERS AGES 15 AND OLDER			
1. ID CODE	12. What is [NAME]'s main activity or employment status?	13. What is the nature of [NAME]'s current agricultural activity?	14. Has [NAME] obtained any formal training in agriculture?
	GOVERNMENT..... SEMI-GOV'T..... PRIVATE SECTOR..... EMPLOYER (AGRI)..... EMPLOYER (NON-AGRI)..... OWN ACCOUNT (AGRI)..... OWN ACCOUNT (NON-AGRI)..... CONTRIBUTING FAMILY MEMBER (AGRI).....		YES..... 1 NO..... 2 ▶ NEXT HH MEMBER
HH01	13	1	



2. Data Validation Error

ONLY FOR HOUSEHOLD MEMBERS AGES 15 AND OLDER			
1. ID CODE	12. What is [NAME]'s main activity or employment status?	13. What is the nature of [NAME]'s current agricultural activity?	14. Has [NAME] obtained any formal training in agriculture?
	GOVERNMENT..... SEMI-GOV'T..... PRIVATE SECTOR..... EMPLOYER (AGRI)..... EMPLOYER (NON-AGRI)..... OWN ACCOUNT (AGRI)..... OWN ACCOUNT (NON-AGRI)..... CONTRIBUTING FAMILY MEMBER (AGRI).....		YES..... 1 NO..... 2 ▶ NEXT HH MEMBER
HH01	16		



3. Missing Error

ONLY FOR HOUSEHOLD MEMBERS AGES 15 AND OLDER			
1. ID CODE	12. What is [NAME]'s main activity or employment status?	13. What is the nature of [NAME]'s current agricultural activity?	14. Has [NAME] obtained any formal training in agriculture?
	GOVERNMENT..... SEMI-GOV'T..... PRIVATE SECTOR..... EMPLOYER (AGRI)..... EMPLOYER (NON-AGRI)..... OWN ACCOUNT (AGRI)..... OWN ACCOUNT (NON-AGRI)..... CONTRIBUTING FAMILY MEMBER (AGRI).....		YES..... 1 NO..... 2 ▶ NEXT HH MEMBER
HH01		3	



4. Logical Error

ONLY FOR HOUSEHOLD MEMBERS AGES 15 AND OLDER			
1. ID CODE	12. What is [NAME]'s main activity or employment status? GOVERNMENT..... SEMI-GOV'T..... PRIVATE SECTOR.. EMPLOYER (AGRI) EMPLOYER (NON- AGRI)..... OWN ACCOUNT (AG OWN ACCOUNT (NO AGRI)..... CONTRIBUTING FAMILY MEMBE (AGRI)..... 8	13. What is the nature of [NAME]'s current agricultural activity?	14. Has [NAME] obtained any formal training in agriculture? YES..... 1 NO..... 2 ▶ NEXT HH MEMBER
HH01	6	3	



Empirical Strategy

Skip Error Model: $A_{ijc} = \alpha + \beta \times CAPI_i + \delta \times HH + \epsilon$

Missing Error Model: $B_{ijc} = \alpha + \beta \times CAPI_i + \delta \times HH + \epsilon$

Validation Error Model: $C_{ijc} = \alpha + \beta \times CAPI_i + \delta \times HH + \epsilon$

Logic Error Model: $D_{ijc} = \alpha + \beta \times CAPI_i + \delta \times HH + \epsilon$

Total Error Model: $Z_{ijc} = \alpha + \beta \times CAPI_i + \delta \times HH + \epsilon$

where,

$CAPI_i$ is an indicator variable for CAPI (1 if CAPI)

HH is a matrix of household characteristics (age, sex, etc.)

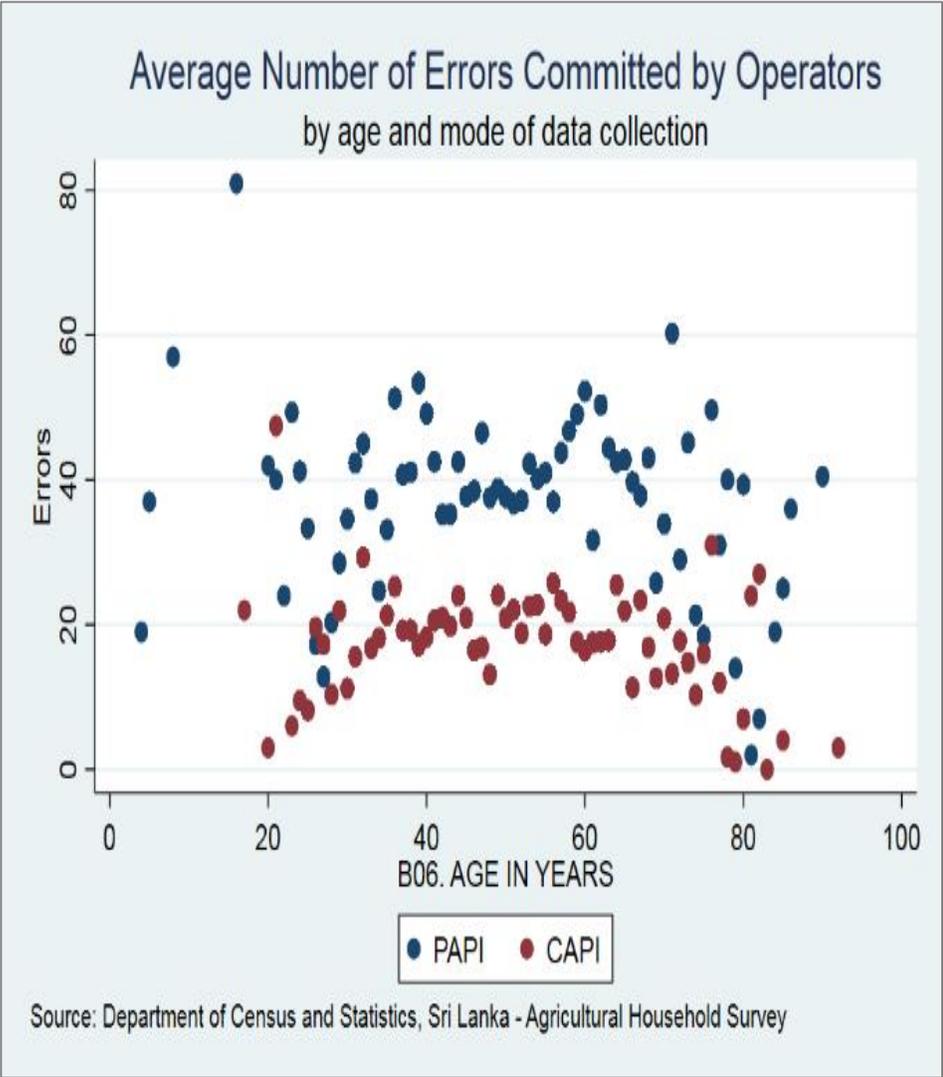
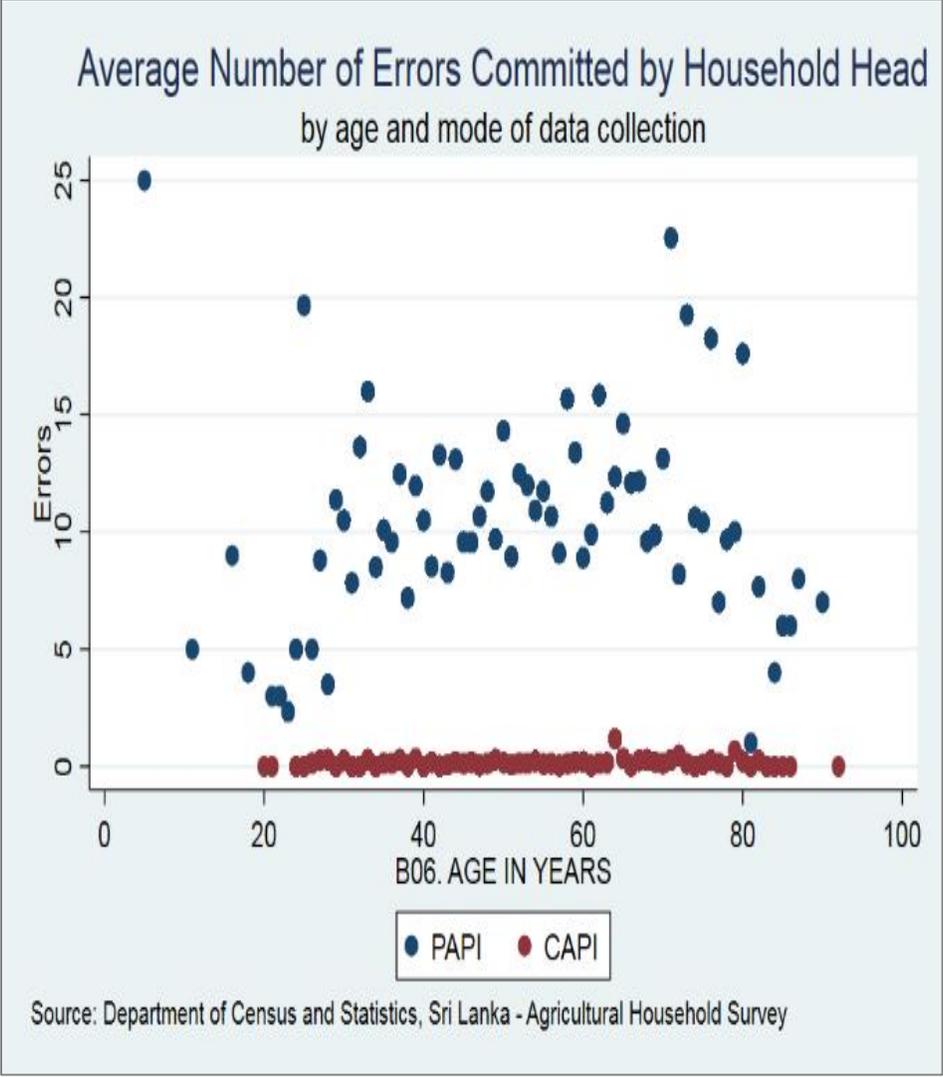
Table 2. Survey Error Models

VARIABLES	(1) Total Error Model	(2) Skip Model	(3) Validation Model	(4) Missing Model	(5) Logic Model
CAPI (base=PAPI)	-37.37*** (10.11)	-8.803*** (1.052)	-1.720*** (0.629)	-26.52*** (9.380)	-0.323** (0.150)
Sex (base=Male)	-6.091*** (1.485)	-0.803** (0.325)	-0.378** (0.187)	-4.822*** (1.269)	-0.0877 (0.0563)
Age (in years)	0.0807* (0.0464)	0.0113 (0.0111)	0.00102 (0.00688)	0.0676* (0.0390)	0.000817 (0.00208)
Quartile of Assets (base=Q1)					
Q2	4.607*** (1.273)	0.351 (0.311)	0.357** (0.165)	3.840*** (1.062)	0.0592 (0.0644)
Q3	8.387*** (2.031)	0.320 (0.460)	0.950*** (0.260)	6.939*** (1.759)	0.178 (0.112)
Q4	10.35*** (1.611)	1.150*** (0.384)	0.774*** (0.217)	8.318*** (1.344)	0.105 (0.0864)
Education (base=Secondary and below)	-1.476 (1.326)	0.381 (0.296)	-0.231 (0.163)	-1.637 (1.128)	0.0104 (0.0629)
Main Activity (base=Others)	6.142*** (1.118)	0.505* (0.296)	0.493*** (0.146)	5.145*** (0.939)	-0.00215 (0.0462)
Number of Operator/s	11.28** (4.599)	1.528 (1.021)	0.470 (0.317)	9.153** (3.893)	0.131 (0.144)
Constant	18.93* (11.09)	5.801*** (1.635)	0.970 (0.797)	12.08 (10.08)	0.0758 (0.238)
Observations	3,074	3,074	3,074	3,074	3,074
R-squared	0.481	0.508	0.409	0.469	0.434
EA FE	Yes	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Sampling weights	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

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

Figure 1: Error patterns by household head versus operator



Q2. What are the cost implications of switching to CAPI?

Empirical Strategy

Cost Comparison Function

$$FC_{CAPI} + (VC_{CAPI} \times \# \text{ of Questionnaires}) = FC_{PAPI} + (VC_{PAPI} \times \# \text{ of Questionnaires})$$

Where,

FC denotes fixed cost

VC denotes variable cost

Table 3. Cost Comparison Analysis

Costs	PAPI		CAPI	
	Per Interview	Total <i>For 1,467 interviews</i>	Per Interview	Total <i>For 1,467 interviews</i>
A. Fixed Costs (FC)				
1. Software	Free		Free	
2. Programming	N/A	\$8,000 ^a	N/A	\$16,400 ^a
Subtotal FC		\$8,000		\$16,400
B. Variable Costs (VC)				
1. Tablets (including data and power bank)	N/A		\$0.07 ^b	\$102.69
2. Enumerator	<i>*Not included in the cost analysis as this is the same for both methods</i>			
3. Data entry	\$2.00	\$2,934.00	N/A	
4. Printing	\$1.30	\$1,907.10	N/A	
5. Miscellaneous costs (storage, bags, logistics, etc.)	\$3.00	\$4,401.00	\$0.50	\$733.50
Subtotal VC	\$6.30	\$9,242.10	\$0.57	\$836.19
Total Cost (FC+VC)^c		\$17,242.10		\$17,236.19
Breakeven point at which CAPI becomes more cost-effective than PAPI				
		1,467 interviews		

Q3. How do respondents perceive CAPI?

Table 4. Respondent Perception

Perception Questions	CAPI	PAPI
Q1: What did you think of the duration of the interview?		
Very Short	0.06%	0.17%
Short	7.35%	5.70%
Length is just right	43.85%	49.91%
Long	36.79%	36.85%
Very Long	11.96%	7.36%
Q2: Did you enjoy participating in the interview?		
Yes	76.25%	90.45%
No	23.75%	9.55%
Q3: How satisfied or dissatisfied are you with the interview based on the flow of questions?		
Very Satisfied	1.42%	3.37%
Satisfied	46.53%	69.05%
Neither Dissatisfied nor Satisfied	46.70%	22.73%
Dissatisfied	3.36%	2.57%
Very Dissatisfied	1.99%	2.28%
Q4: Did you find the questions easy or difficult to answer?		
Very Easy	1.59%	5.83%
Easy	53.08%	69.37%
Neither Difficult nor Easy	38.27%	18.63%
Difficult	5.24%	4.69%
Very Difficult	1.82%	1.49%

Q5: If we went through the survey again, do you think any answers would change?		
Yes	8.49%	6.74%
No	91.51%	93.26%
Q6: Did you feel comfortable talking to the interviewer?		
Yes	91.74%	95.42%
No	8.26%	4.58%
Q7: If we're not recording the answers (just talking to you), how much would your answer have changed?		
Totally different	0.34%	1.43%
A bit different	28.47%	17.75%
No change	71.18%	80.81%

Q4. What is the impact of CAPI on key agricultural statistics such as rice yield?

Table 5. Impact on Policy Analysis

VARIABLES	Yield
CAPI (base=PAPI)	-1.773** (0.893)
Harvested Area (in hectares)	-0.128**** (0.0403)
Female (base=Male)	0.729 (1.400)
Age	0.127 (0.115)
Squared Age	-0.00130 (0.00114)
Marital Status (base=Never Married)	
Married	1.103** (0.527)
Widowed	0.717 (1.155)
Divorced	0.0981 (1.003)
Separated	0.0378 (1.102)
Above Secondary (base=Secondary and below)	-0.254 (0.557)
Agri Training Yes (base=No)	-0.179 (0.587)
Constant	1.054 (2.597)
Observations	548
R-squared	0.187
EA FE	Yes
Equip FE	Yes
Sampling weights	Yes

Robust standard errors in parentheses
 **** p<0.01, ** p<0.05, * p<0.1

- The average yield in CAPI is 1.7 tons/hectare lower when compared to PAPI.

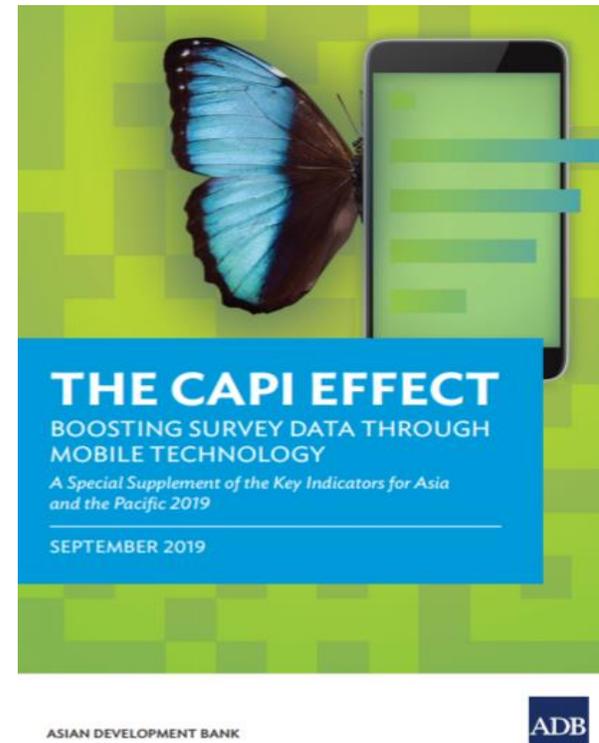
Conclusion

The following conclusions are drawn from the study:

- The results corroborate the literature in terms of the perceived benefits of CAPI for data quality where CAPI seems to reduce the number of errors committed per questionnaire.
- Cost analysis showed that costs were in favor of CAPI for medium to large scale surveys of more than 1467 households.
- There are noticeable differences in perception between CAPI and PAPI. Sensitizing the population before conduct of a survey on CAPI should allay any fear of technology.
- These differences in data quality may have significant implications for policy analysis.

For more information!

[The CAPI Effect: Boosting Survey Data Through Mobile Technology](#)



For more information!

<https://development.asia/explainer/how-mobile-technology-enables-better-sdgs-monitoring>

The screenshot shows the top navigation bar of the Development Asia website. The logo 'DEVELOPMENT ASIA' is on the left, with the tagline 'An initiative of Asian Development Bank'. To the right, there are social media icons for Facebook, Twitter, and LinkedIn, and a search bar. Below the navigation bar, a red 'EXPLAINER' tag is positioned above the article title 'How Mobile Technology Enables Better SDGs Monitoring'. The main image shows a group of people sitting on the ground outdoors, with one person using a smartphone. Below the image is a caption: 'Staff from the Ministry of Agricultural and Rural Development in Viet Nam conduct a survey among farmers using smartphones. Photo credit: ADB.' Below the image are social sharing icons for Facebook, LinkedIn, and Twitter, along with icons for email, download, and print. To the right of the image is the 'Ask the Experts' section, featuring two experts: Lakshman Nagraj Rao, Statistician, Economic Research and Regional Cooperation Department, Asian Development Bank; and Pamela Lapitan, Associate Statistics Officer, Economic Research and Regional Cooperation Department, Asian Development Bank. Below the image is the 'Introduction' section, which discusses the importance of accurate and timely data for achieving the Sustainable Development Goals (SDGs).

DEVELOPMENT ASIA
An initiative of Asian Development Bank

Finding Solutions Together

Follow us

Search our content.

[Home](#) [About](#) [Topics](#) [Countries](#) [Sustainable Development Goals](#) [All content](#) [Communities](#)

EXPLAINER

How Mobile Technology Enables Better SDGs Monitoring

Staff from the Ministry of Agricultural and Rural Development in Viet Nam conduct a survey among farmers using smartphones. Photo credit: ADB.

Share on:

Ask the Experts

Lakshman Nagraj Rao
Statistician, Economic Research and Regional Cooperation Department, Asian Development Bank

Pamela Lapitan
Associate Statistics Officer, Economic Research and Regional Cooperation Department, Asian Development Bank

Introduction

Achieving the Sustainable Development Goals (SDGs) depends on accurate, timely, consistent, and comparable data. Countries need to keep track of progress toward 169 targets and 232 performance indicators. This is an arduous task especially for national statistics offices (NSOs), which play a key role in collecting, validating, and reporting data. Several indicators rely on census and survey data that are costly to collect and take a long time to process.