

# Assessing input use through SMS surveys: Evidence from Tanzania

Beliyou Haile, Jeffrey Dickinson  
Jawoo Koo, Carlo Azzarri

Boston, USA  
August 3 - 4, 2016

- Agriculture in sub-Saharan Africa
  - Low cereal yields and adoption of improved technologies, recycling of hybrids (Gollin et al., 2005; Westengen et al, 2014)
  - Low purchasing power, limited market access, information barrier (Foster and Mark, 2014, 2012; Duflo et al., 2008; Suri, 2011)
  - Expansion of mobile phone infrastructure, ICT for agricultural and economic development (Furuholt and Matotay, 2011)
  - E.g., World Bank's Listening to Africa (L2A) pilot program

# Research question and hypothesis

- **Question:** What is the linkage between market access and hybrid maize seed use/purchase?
- **Hypothesis:** Farmers with limited market access are less likely to use hybrids/more likely to recycle seeds
- **Case study:** Tanzania – IFPRI/PIM-GeoPoll Short Message Service (SMS) survey

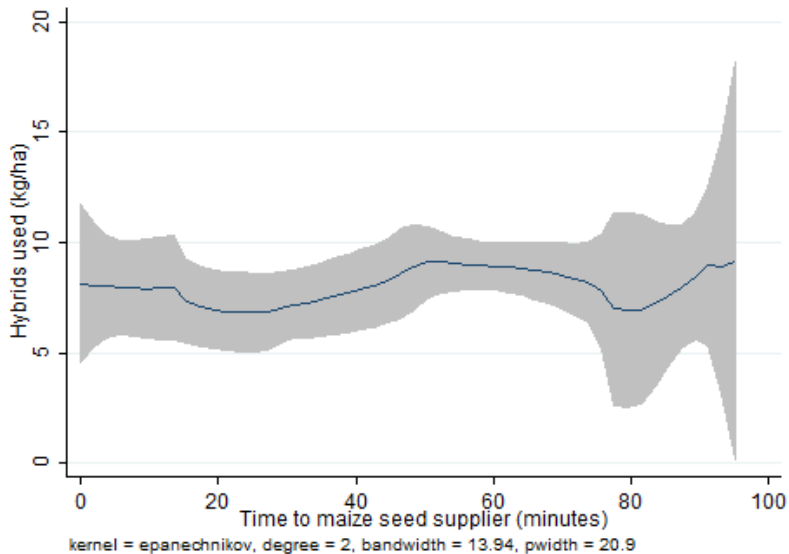
- Two waves: 2014 and 2015
- *Ex-ante* sample of 1,000 individuals (2014), allocated into across regions using Neyman criterion

	2014		2015	
	Count	Percent	Count	Percent
Surveys sent	4,993	100%	3,407	100%
Opt ins	1,464	29%	1,552	46%
Completes	1,001	20%	1000	29%
Ineligible	39	1%	464	14%
Refusals	35	1%	88	3%
Non-response	2,454	49%	1,767	51.86%
<i>No. of questions</i>	<i>16</i>		<i>18</i>	
<i>Minutes (average)</i>	<i>52</i>		<i>61</i>	

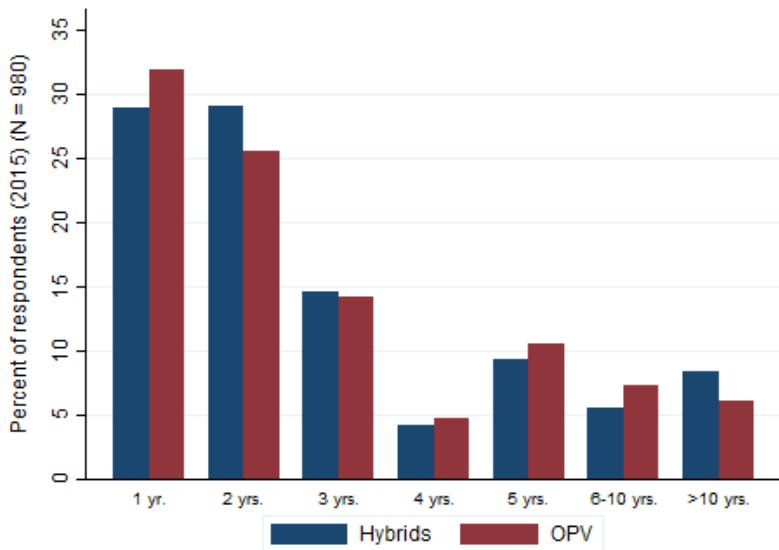
- Two waves: 2014 and 2015
- *Ex-ante* sample of 1,000 individuals (2014), allocated into across regions using Neyman criterion

	2014		2015	
	Count	Percent	Count	Percent
Surveys sent	4,993	100%	3,407	100%
Opt ins	1,464	29%	1,552	46%
Completes	1,001	20%	1000	29%
Ineligible	39	1%	464	14%
Refusals	35	1%	88	3%
Non-response	2,454	49%	1,767	51.86%
<i>No. of questions</i>	16		18	
<i>Minutes (average)</i>	52		61	

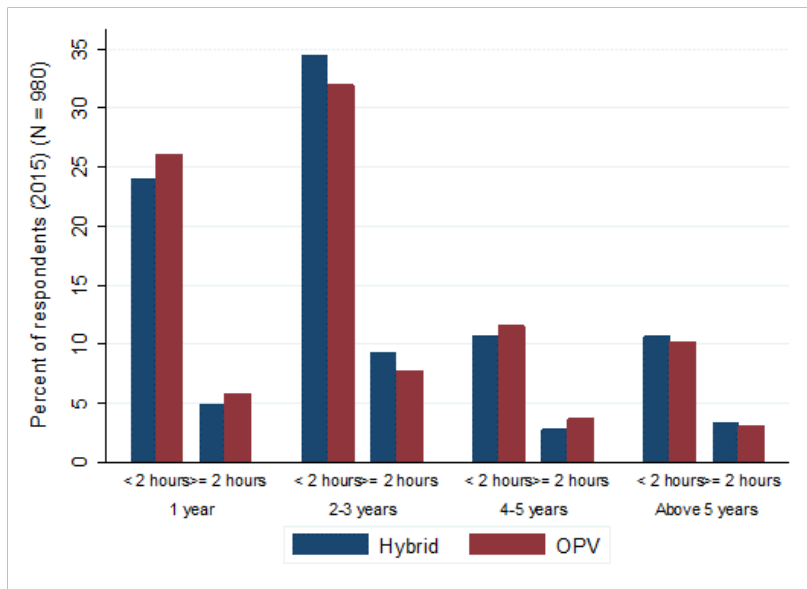
## Hybrids used (kg/ha) by travel time



## Maize seed recycling (by type)/1

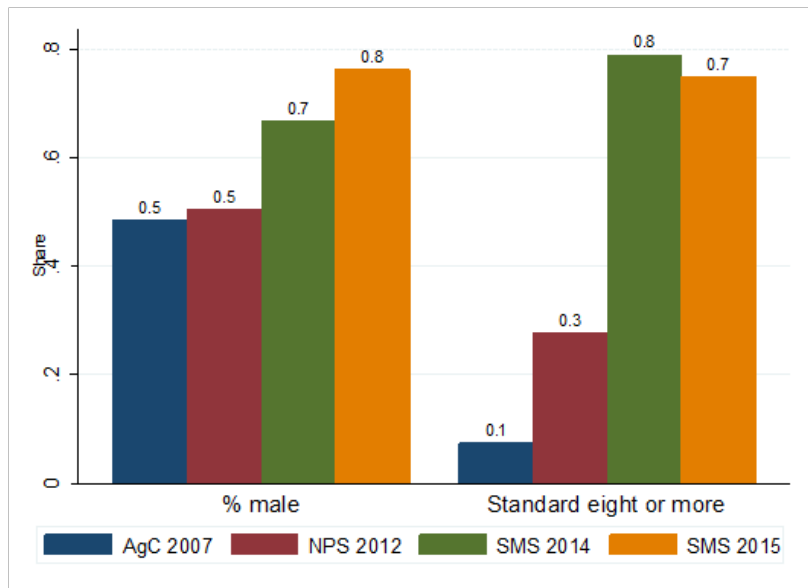


## Maize seed recycling (by travel time and type)/2

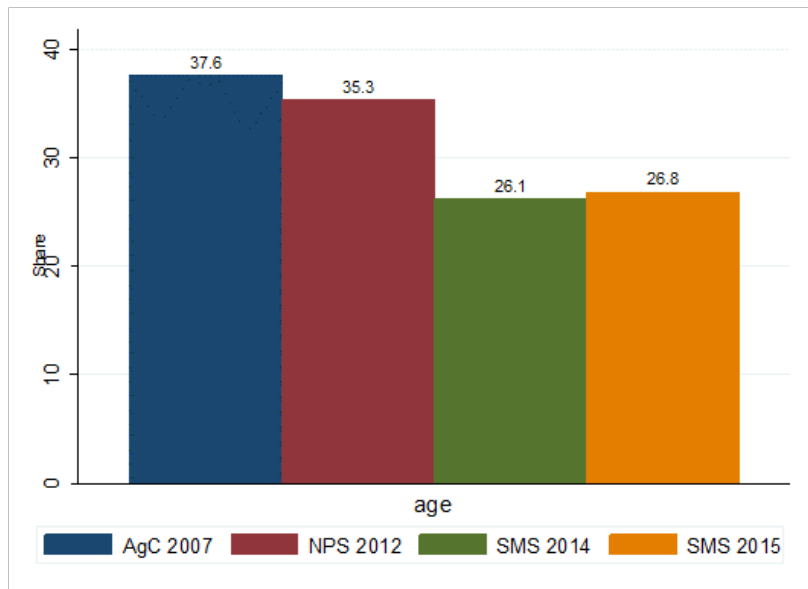




## Comparability of SMS sample with NPS 2012/1



## Comparability of SMS sample with NPS 2012/2



# Models and sample adjustment

$$Y_i = \alpha_0 + \alpha_1 \text{Travel2supplier}_i + \alpha' \mathbf{X}_i + \mathbf{r} + v_i \quad (1)$$

$$Y_{it} = \beta_0 + \beta_1 \text{Travel2supplier}_{it} + \beta' \mathbf{X}_{it} + \mathbf{r} + \mu_i + \epsilon_{it} \quad (2)$$

$$w_{i(SMS)}^g = Sh_{NPS}^g / Sh_{SMS}^g \quad (3)$$

- $Y$  indicator for seed purchase, quantity of hybrids used–kg/ha)
- *Travel2supplier* indicator for travelling  $\geq 2$  hours
- $w_i$  = adjustment weight,  $Sh$  = share of group  $g$  (age, gender, education, or a composite index)
- Over-represented (under-represented)  $g$  in SMS, relative to the distribution of  $g$  in NPS, are associated lower (higher) weight

**Table 1:** Likelihood of seed purchase in the last season (**wave 1**) (logit)

	1	2	3	4	5
	$\partial y/\partial x$	$\partial y/\partial x$	$\partial y/\partial x$	$\partial y/\partial x$	$\partial y/\partial x$
<b>Travels more than two hours</b>	0.011	0.014	0.042	-0.012	0.021
	(0.048)	(0.071)	(0.053)	(0.070)	(0.086)
Age	0.004*		0.003	-0.001	
	(0.002)		(0.003)	(0.003)	
Male	0.069*	0.134**		0.041	
	(0.036)	(0.054)		(0.054)	
Education==Standard eight or more	0.030	0.042	0.011		
	(0.043)	(0.059)	(0.047)		
Price of OPV maize seeds ('000 TSH/kg)	0.002	-0.003	0.002	-0.002	-0.001
	(0.003)	(0.004)	(0.003)	(0.004)	(0.006)
Price of hybrid maize seeds ('000 TSH/kg)	0.004	0.012*	0.005	0.003	0.008
	(0.005)	(0.007)	(0.005)	(0.007)	(0.008)
Total land area (ha)	-0.015*	-0.022*	-0.017**	-0.013	-0.010
	(0.008)	(0.011)	(0.008)	(0.011)	(0.013)
Maize suitability index	0.001**	0.002***	0.001**	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Number of observations			958		
Log-Likelihood	-642.01	-777.56	-670.60	-606.90	-309.45
chi2	34.531	22.456	25.375	5.380	2.954
p	0.000	0.008	0.003	0.800	0.889

\* $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Robust standard errors in parentheses. Zone fixed effects included  
 1=Unweighted, 2=age-based weight (wt), 3=gender-based wt  
 4=education-based wt, 5=composite wt

Table 2: Hybrids used/recycled in the last season (kg/ha) (**wave 1**) (OLS)

	1	2	3	4	5
	coef/se	coef/se	coef/se	coef/se	coef/se
<b>Travels more than two hours</b>	2.390** (1.004)	2.155 (1.710)	2.226** (1.076)	5.083*** (1.841)	5.375** (2.104)
Age	0.037 (0.043)		0.049 (0.048)	0.105 (0.069)	
Male	0.309 (0.711)	0.245 (0.952)		0.256 (1.120)	
Education==Standard eight or more	-1.247 (0.791)	-0.944 (1.052)	-1.229 (0.844)		
Most recent season ==long rains	-0.688 (0.695)	-1.174 (1.004)	-0.311 (0.712)	-0.007 (1.035)	-0.532 (1.298)
Price of OPV maize seeds ('000 TSH/kg)	-0.085 (0.084)	-0.048 (0.108)	-0.056 (0.079)	-0.027 (0.134)	-0.052 (0.124)
Price of hybrid maize seeds ('000 TSH/kg)	-0.756*** (0.086)	-0.809*** (0.112)	-0.724*** (0.087)	-0.811*** (0.136)	-0.673*** (0.153)
Total land area (ha)	-0.805*** (0.108)	-0.667*** (0.114)	-0.823*** (0.116)	-0.497*** (0.178)	-0.594*** (0.223)
Maize suitability index	-0.001 (0.011)	0.002 (0.014)	-0.013 (0.013)	-0.012 (0.021)	-0.037 (0.026)
R2	0.124	0.130	0.126	0.143	0.155

\*p<0.1, \*\* p<0.05, \*\*\* p<0.01. Zone fixed effects included

1=Unweighted, 2=age-based weight (wt), 3=gender-based wt, 4=education-based wt, 5=composite wt

**Table 3: Hybrids used/recycled in the last season (kg/ha) (wave 2) (OLS)**

	1	2	3	4	5
	coef/se	coef/se	coef/se	coef/se	coef/se
<b>Travels more than two hours</b>	0.043 (0.984)	1.392 (1.366)	0.615 (1.108)	-0.822 (1.551)	-1.444 (2.203)
Age	0.003 (0.053)		-0.009 (0.062)	-0.077 (0.060)	
Male	0.496 (0.943)	0.457 (1.377)		0.610 (1.720)	
Education==Standard eight or more	-1.825* (1.092)	-0.909 (1.416)	-1.873 (1.262)		
Most recent season ==long rains	-1.648* (0.894)	-1.370 (1.368)	-1.973** (0.996)	-3.672** (1.484)	-6.681*** (2.096)
Price of OPV maize seeds ('000 TSH/kg)	-0.055 (0.090)	-0.093 (0.103)	-0.031 (0.111)	-0.083 (0.116)	0.053 (0.182)
Price of hybrid maize seeds ('000 TSH/kg)	-0.743*** (0.113)	-0.747*** (0.196)	-0.765*** (0.123)	-0.774*** (0.177)	-1.039*** (0.183)
Total land area (ha)	-0.551*** (0.142)	-0.491** (0.197)	-0.453*** (0.151)	-0.583** (0.246)	-1.061*** (0.286)
Maize suitability index	0.003 (0.016)	0.001 (0.028)	0.010 (0.017)	0.014 (0.027)	0.011 (0.029)
R2	0.093	0.084	0.093	0.110	0.173

\*p<0.1, \*\* p<0.05, \*\*\* p<0.01. Zone fixed effects included

1=Unweighted, 2=age-based weight (wt), 3=gender-based wt, 4=education-based wt, 5=composite wt

**Table 4: Hybrids used/recycled in the last season (kg/ha) (panel) (random effects)**

	1 coef/se	2 coef/se	3 coef/se	4 coef/se	5 coef/se
<b>Travels more than two hours</b>	2.557** (1.059)	1.856 (1.387)	2.856** (1.139)	2.513 (1.548)	1.655 (1.929)
Age	0.039 (0.051)		0.016 (0.057)	0.025 (0.074)	
Male	0.812 (0.898)	1.104 (1.085)		0.166 (1.610)	
Education==Standard eight or more	-2.364** (1.183)	-0.675 (1.338)	-2.372* (1.362)		
Most recent season ==long rains	0.050 (0.878)	0.486 (1.301)	0.331 (0.917)	0.076 (1.287)	-1.771 (1.800)
Price of OPV maize seeds ('000 TSH/kg)	0.113 (0.114)	0.136 (0.117)	0.140 (0.133)	0.026 (0.116)	0.165 (0.190)
Price of hybrid maize seeds ('000 TSH/kg)	-0.912*** (0.115)	-0.893*** (0.145)	-0.882*** (0.130)	-1.092*** (0.195)	-1.054*** (0.279)
Total land area (ha)	-0.692*** (0.139)	-0.662*** (0.163)	-0.594*** (0.149)	-0.665*** (0.225)	-0.551** (0.244)
Maize suitability index	-0.013 (0.013)	-0.012 (0.017)	-0.007 (0.014)	-0.024 (0.022)	-0.029 (0.028)
chi2	121.461	70.879	91.407	62.798	41.774
p	0.000	0.000	0.000	0.000	0.000

\* $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Zone fixed effects included

1=Unweighted, 2=age-based weight (wt), 3=gender-based wt, 4=education-based wt, 5=composite wt

# Conclusion

- Expansion of mobile network infrastructure opens up new opportunities for almost real time and less costly data collection
- SMS surveys to complement, validate, and refine other methods (e.g., expert opinion elicitation, household survey)
- Recycling of hybrids quite widespread (as common as recycling of OPV)
- Suggestive evidence of positive association between travel time and use of (recycled) hybrid maize seeds
- However, the association is not consistent across waves and different SMS sampling adjustment weights



- How to improve the (internal and external) validity of SMS survey results?
  - Better understanding of the database of mobile phone subscribers (e.g., age, gender, welfare).
  - Stratify sample to mimic the distribution of target groups in the reference universe drawn from known and representative statistics
  - Reduce unit non-response rate (more frequent reminders? Incentives? Other? without introducing moral hazard and negative self-selection)
    - 57% attrition rate
    - 54 % - 70 % passive non-response rate

- Minimize response and measurement errors (beyond the 160 characters limit...is that feasible?)
  - Varietal specificity (through expert opinion)
  - Local versus metric units
  - Consistency checks
  - Compare SMS survey data with nationally representative surveys (e.g., NPS and Ag Census)
- Geo-referencing (cell tower locations?)

*THANK YOU!*

# References