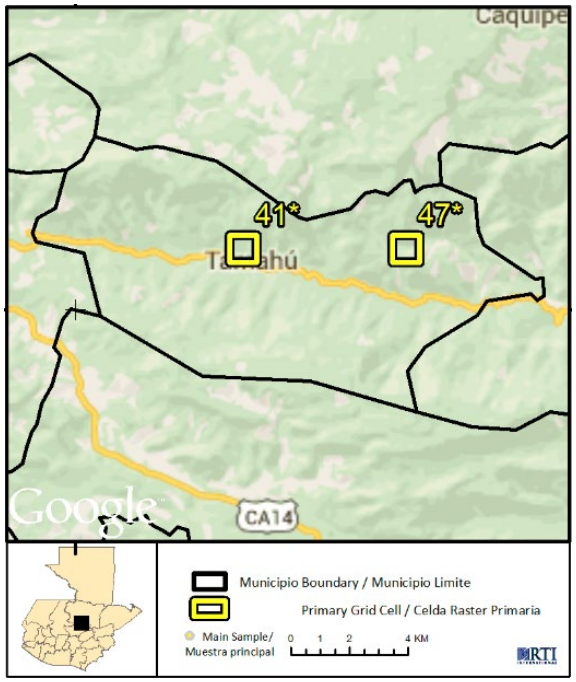
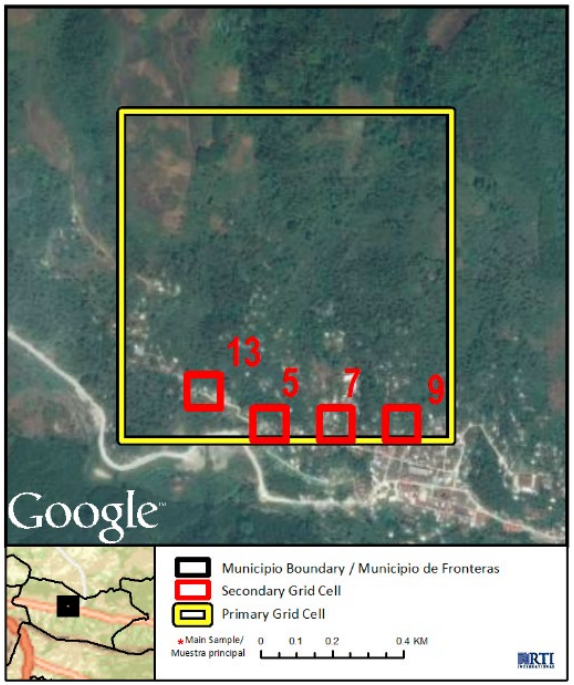
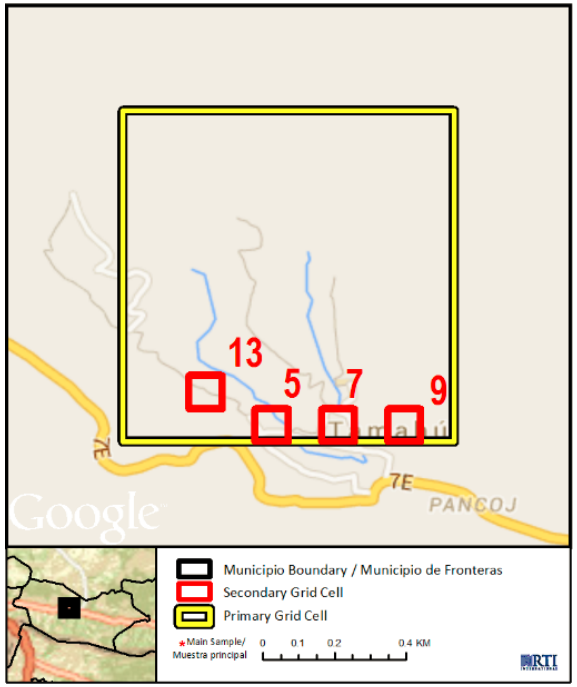
Design Effects in a Geo-sampling Design

3MC, July 2016

Jennifer Unangst
Jeniffer Iriondo-Perez
Dr. Safaa Amer
RTI International







Purpose

What are the impacts of using a geo-sampling design on the precision of study outcomes?

The Study

Study: Barriers to internet adoption

Where: Low/middle income countries

Outcomes:

- Internet use in past 30 days
- Number mobile phones used in last 30 days

The Study

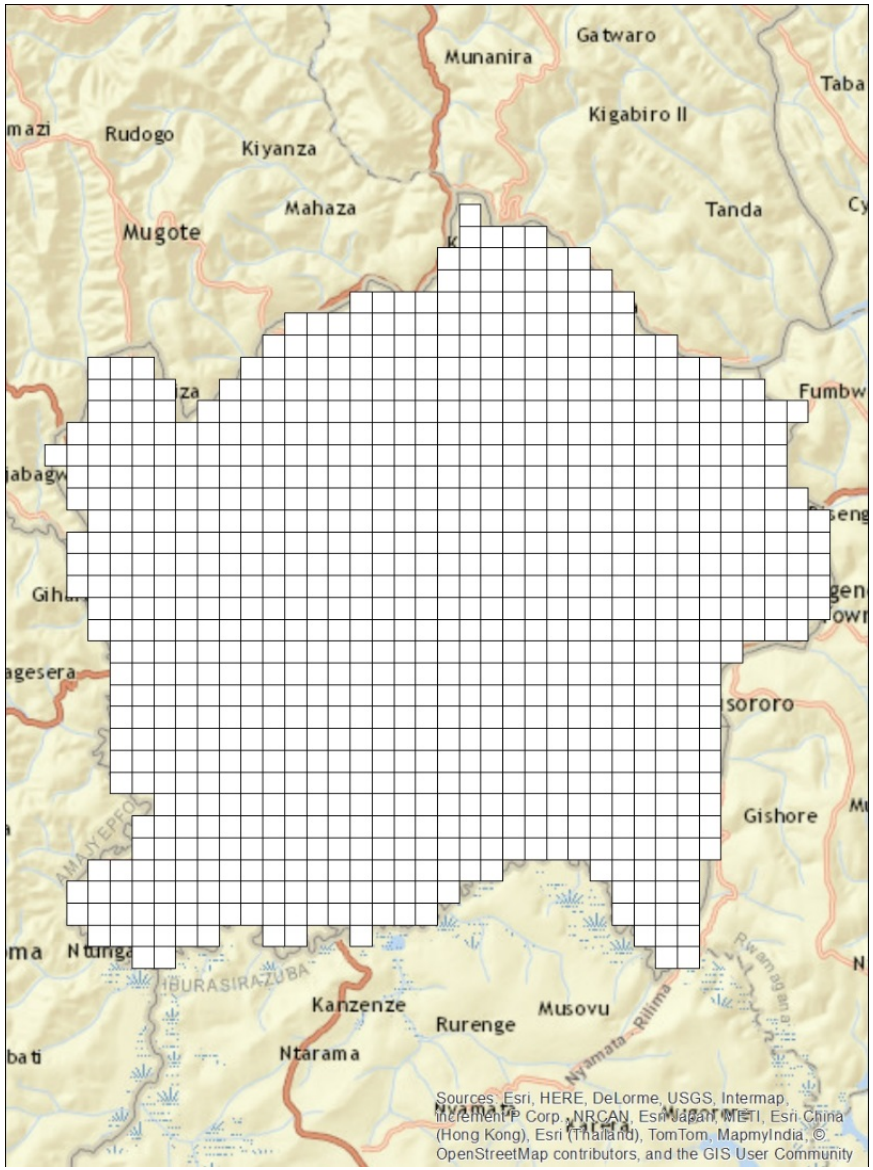
Key Parameters:

- Nationally representative probability-based sample of 15 to 64 year-olds
- 3,000 respondents
- Urban/rural and various poverty levels
- Logistically feasible
- Time and cost effective
  - No listing process
- No random walk
- Adaptable across countries
- Electronic data collection

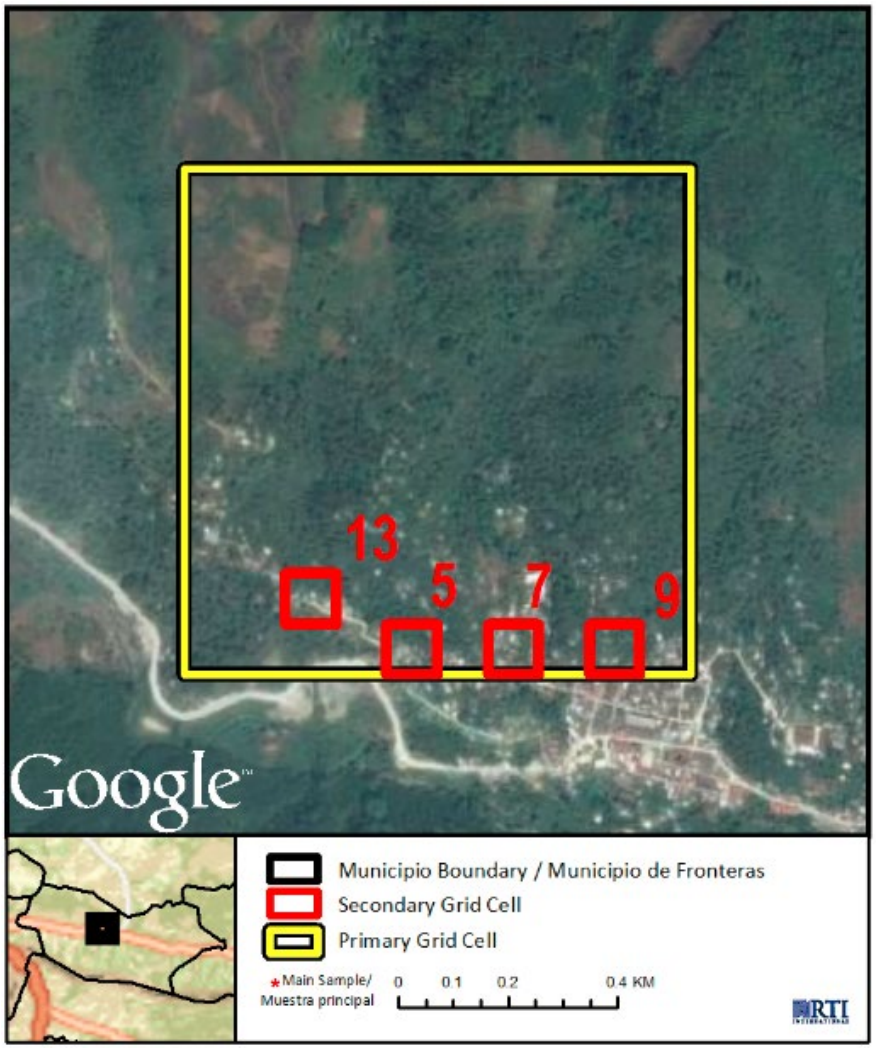
Geo-sampling Methodology

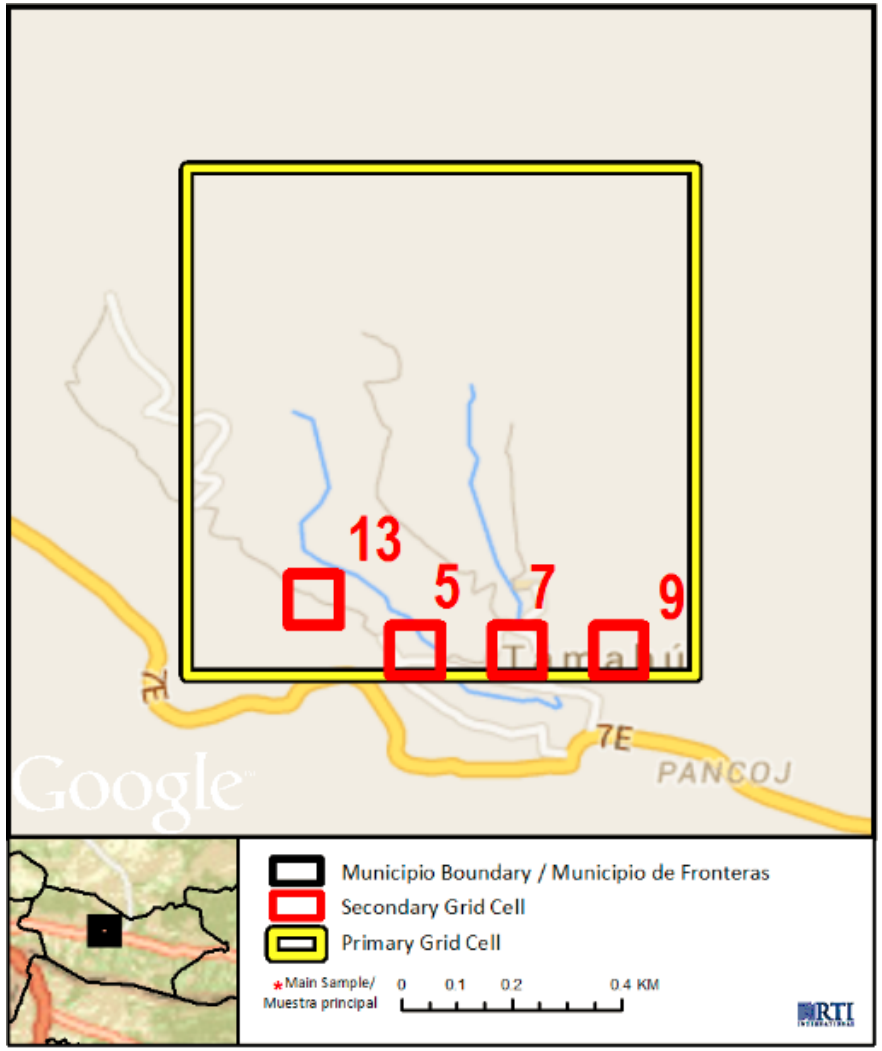
Geo-sampling is:

- An innovative, multi-stage, probability-based sampling methodology
- Integrates Geographic Information Systems (GIS) at several stages of the sampling process
- Divides target areas into logistically manageable units (Amer, AAPOR 2015)



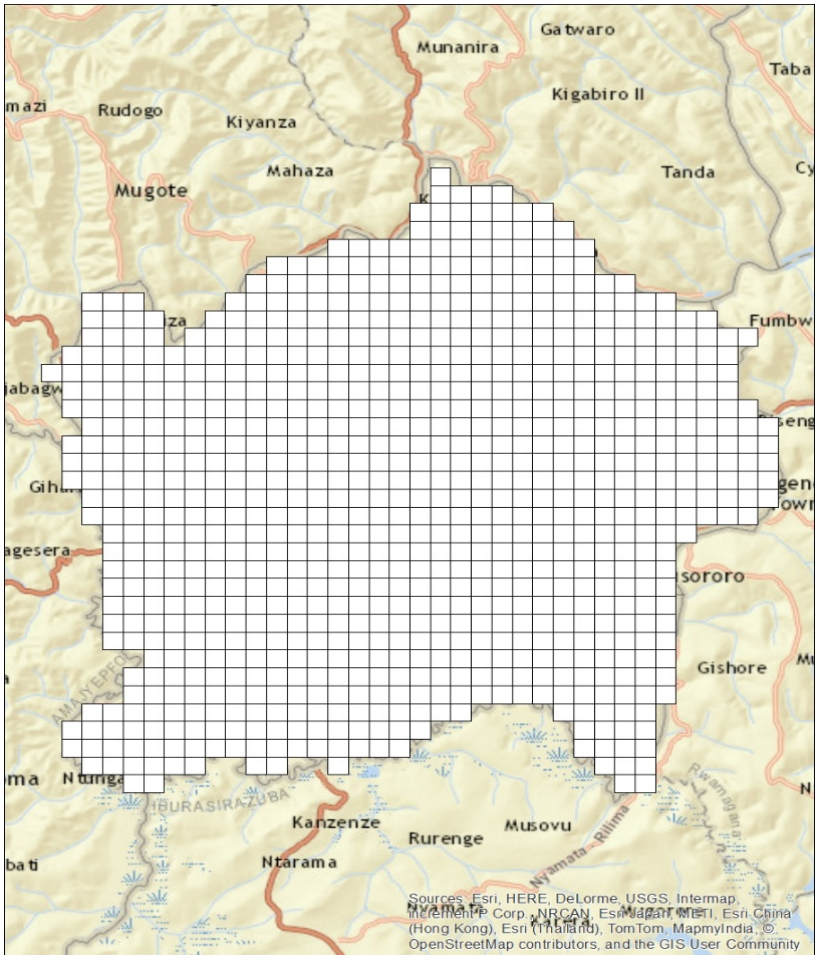
Geo-sampling Methodolodgy





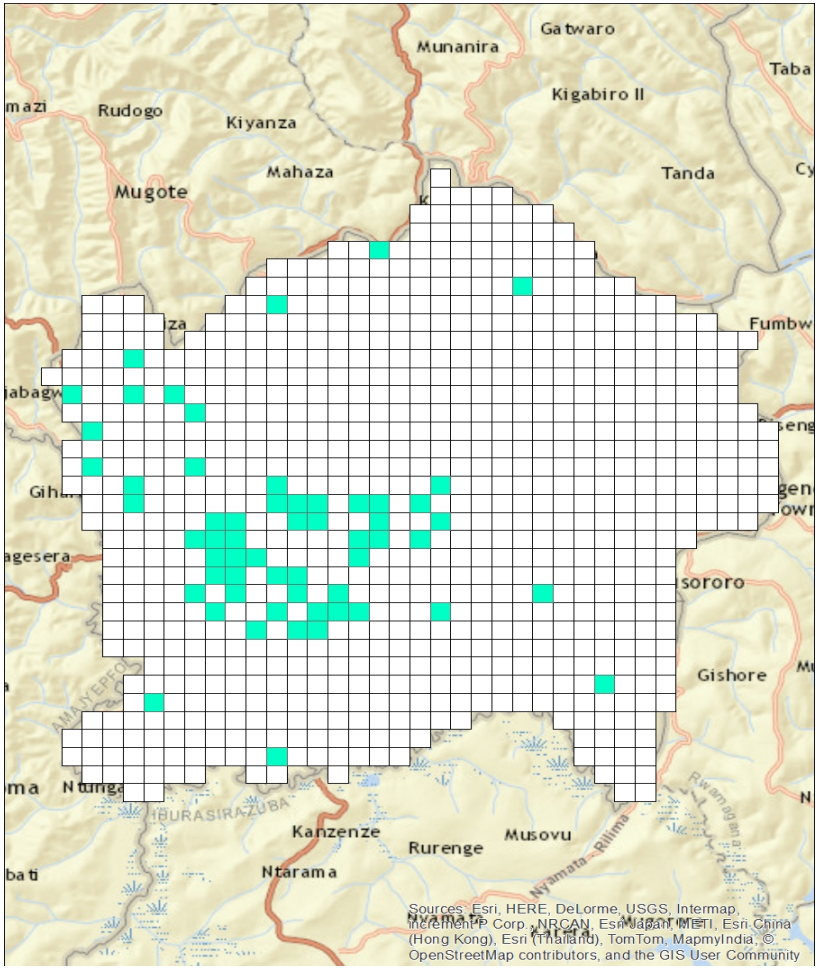
Standardized Steps of Geo-sampling

- Select sample up to “district” level
  - “District” varies by country



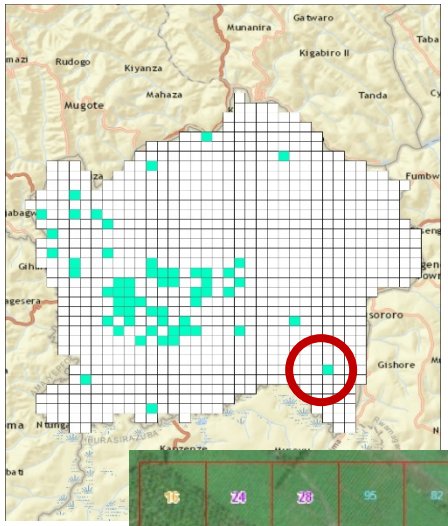
Standardized Steps of Geo-sampling

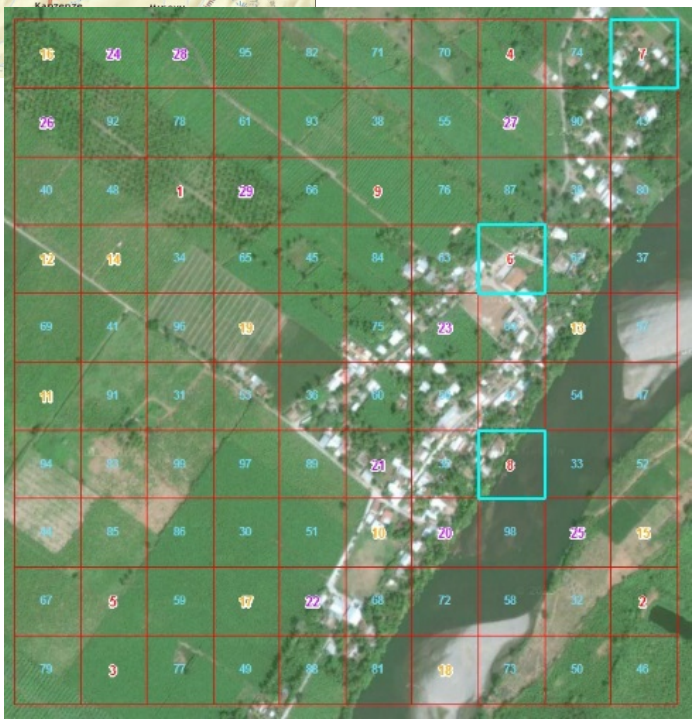
- Select sample up to “district” level
  - “District” varies by country
- Divide districts into 1-km^2 grid cells; (“primary grid units”)
  - Randomly sample residential PGUs



Standardized Steps of Geo-sampling

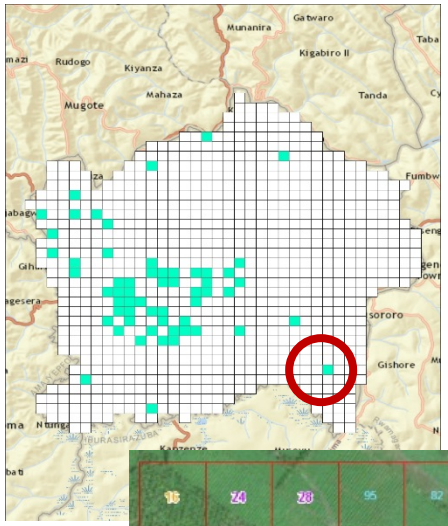
- Select sample up to “district” level
  - “District” varies by country
- Divide districts into 1-km^2 grid cells (“primary grid units”)
  - Randomly sample residential PGUs
- Divide selected 1-km^2 grid cells into smaller sub-grid cells (“secondary grid units”)
  - Randomly sample residential SGUs
  - 100 m^2 in rural areas / 50 m^2 in urban areas

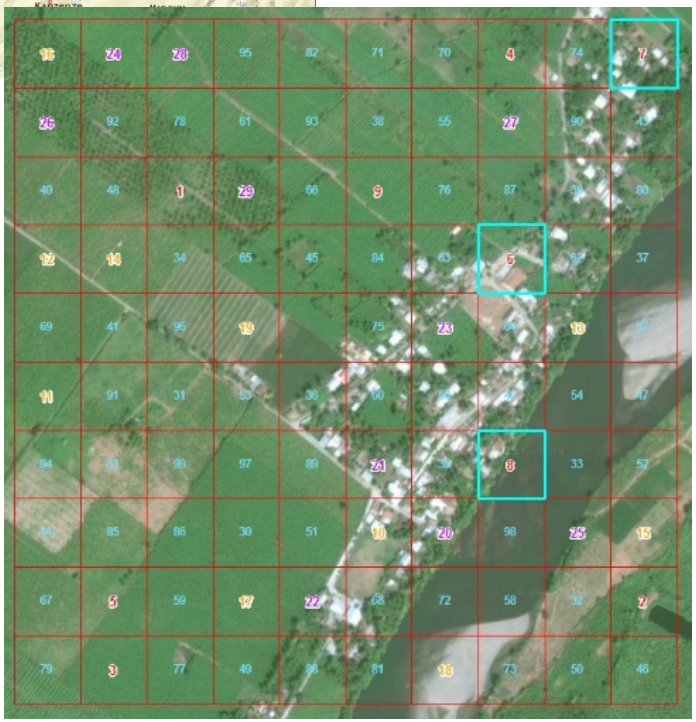




Standardized Steps of Geo-sampling

- Select sample up to “district” level
  - “District” varies by country
- Divide districts into 1-km^2 grid cells (“primary grid units”)
  - Randomly sample residential PGUs
- Divide selected 1-km^2 grid cells into smaller sub-grid cells (“secondary grid units”)
  - Randomly sample residential SGUs
  - 100 m^2 in rural areas / 50 m^2 in urban areas
- All households within selected SGUs
- One adult 15-64 per household





Standardization/Tailoring of Sample Design

- **Countries**
  - **Asia**: Thailand, Bangladesh, India
  - **Africa:** Ghana, Kenya, Nigeria, Rwanda, Uganda
  - **South America:** Brazil, Colombia, Guatemala

- **Variability in sample design resources across countries**
  - Different geographic boundaries (“district”)
  - Availability of administrative unit frames
  - Availability/level of Census data for stratification

- **Country-specific tailoring**
  - Number of stages prior to geo-sampling
  - Size of area units selected before grid units
  - Sampling method (e.g., SRS, probability proportional to size)

*Most tailoring occurred at higher stages before geo-sampling stages

Purpose

What are the impacts of using a geo-sampling design on the precision of study outcomes?

Design Effect

Design Effect (DEFF)

- "Penalty" to the precision of study outcomes from using a complex sample design (e.g., unequal selection probabilities, clustering, stratification) instead of simple random sampling
- Determines effective sample size
- Decomposed conceptually as product of unequal weighting effect (*UWE*) and clustering effects ($DEFF_{clust}$)

$$DEFF = UWE \times DEFF_{clust}$$

Internet Used in Past 30 Days
Design Effects

	Africa		Asia	South America	
	Nigeria	Uganda	Thailand	Brazil	Colombia
Sample Size	3,042	3,071	3,134	3,030	2,704
DEFF	**15.0**	2.9	2.0	8.9	3.8
UWE (% of DEFF)	2.2 (15%)	2.2 (76%)	2.4 (120%)	2.3 (26%)	3.3 (87%)

Number Mobile Phones Used in Past 30 Days
Design Effects

	Africa		Asia	South America	
	Nigeria	Uganda	Thailand	Brazil	Colombia
Sample Size	2,844	2,635	3,111	3,022	2,577
DEFF	**27.0**	4.1	**18.6**	8.3	7.6
UWE (% of DEFF)	2.2 (8%)	2.3 (56%)	2.4 (13%)	2.3 (28%)	3.4 (45%)

Nigeria: A Case Study
Components of the Clustering Effect

- **Average cluster size**
  - Weighted average is appropriate when cluster sizes vary (Holt 1980; Chromy 2014)

- **Intracluster correlation coefficient (ICC)**
  - Measure of homogeneity within clusters
  - Can vary by outcome
  - Our method: ANOVA model with fully nested random effects for clusters and stratification variables as covariates (Ganninger 2010; Delong and Lokhnygina 2014)

Nigeria: A Case Study
Internet Used in Past 30 Days

Sampling Stage	Cluster Unit	ICC	Average Cluster Size	Clustering Effect
1	State	.04	268.4	12.7
2	Local Government Area	.03	62.2	2.8
3	Primary Grid Unit (1-km^2 cell)	.04	29.6	2.3
4	Secondary Grid Unit	.06	14.6	1.9

Nigeria: A Case Study

Number of Mobile Phones Used in Past 30 Days

Sampling Stage	Cluster Unit	ICC	Average Cluster Size	Clustering Effect
1	State	.15	245.8	36.9
2	Local Government Area	.02	57.7	2.1
3	Primary Grid Unit (1-km^2 cell)	.02	27.4	1.6
4	Secondary Grid Unit	.06	13.5	1.7

Conclusions and future analysis

Benefits of geo-sampling methodology

- Innovative design, adaptable across countries
- No listing/No random walk
- Cuts costs and implementation timeline

High design effects not due to geo-sampling methodology

- Large cluster sizes from higher stages of sampling contributing to clustering effects

Future analysis

- Why do the clustering effects vary by outcome?
- Decompose clustering effects across additional countries

Thank you

Jennifer Unangst
Research Statistician
junangst@rti.org

Jeniffer Iriondo-Perez
Research Statistician
perez@rti.org

Safaa Amer, PhD
Senior Research Statistician
samer@rti.org