Satellite-Aided
Household Survey Sampling
in Low- and Middle-Income Settings

Marco J Haenssgen, University of Oxford

ESRC Research Methods Festival 2018
Recent Advances in Rural Health Survey Methodology
Bath, 3 July 2018

APPROACHES TO SURVEY SAMPLING

Approaches to Survey Sampling

Traditional Approaches

In absence of registers as sampling frames, rural household surveys in LMICs are often based on random walk procedures or household listing. For example, health surveys in LMICs:

First-level cluster selection
- Based on census data, proportional to population size

Second-level household selection

- **Household listing and mapping**
  - Select segments and list every inhabited structure
  - Choose starting point and interval

- **Random walk**
  - Drop pin to select starting household
  - Choose interval and follow road, "spin the bottle" at each intersection

Approaches to Survey Sampling

Satellite-Aided Approaches

Satellite-based approaches are gradually emerging in research.

Cluster random sampling
- WAMPLER ET AL. (2013) in Haiti
- Google Earth → ArcMap → Excel → GPS units

Simple random sampling
- ESCAMILLA ET AL. (2014) in Malawi; SHANNON ET AL. (2012) in Lebanon
- GPS → GIS → Google Earth → Digipoint 2 → GIS → Hawth's Tools → GPS

Selection of starting point for random walk
- GALWAY ET AL. (2012) in Iraq; FLYNN ET AL. (2013) in Canada
- GIS → Google Earth / Maps → Printed maps (GALWAY ET AL., 2012)

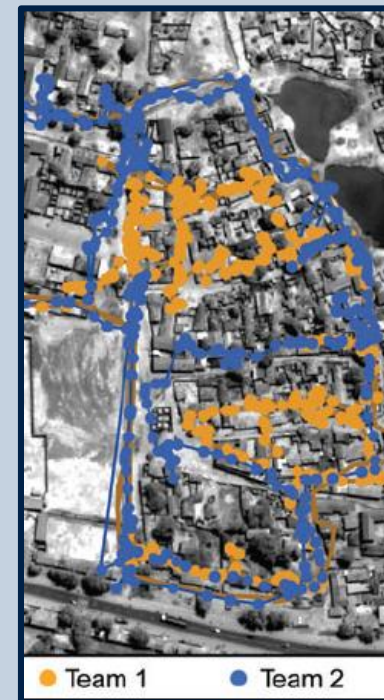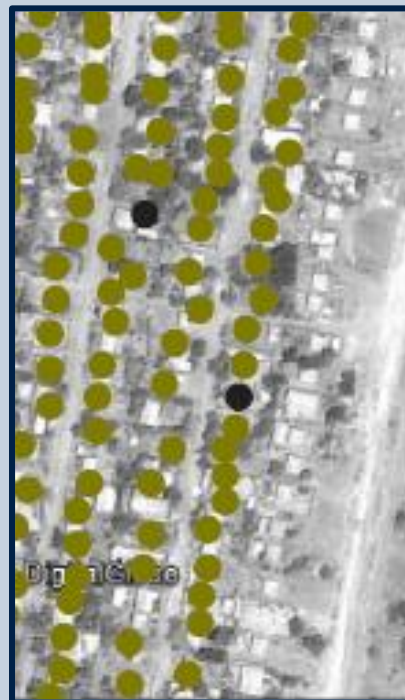**Other (Disease and programme surveillance)**
- CHANG ET AL. (2009), GAMMINO ET AL. (2014), MORLAND & EVENSON (2009)
- GPS → Google Maps / GIS → GPS

Approaches to Survey Sampling

Satellite-Aided Approaches

Researchers have started to incorporate satellite-based approaches in survey research.

Team 1   Team 2

Wampler et al. (2013)
Escamilla et al. (2014)
Galway et al. (2012)
Gammino et al. (2014)

Approaches to Survey Sampling

Issues

Traditional Approaches

- Time and money intensive

- Segmentation to manage workload

- Clustering, high street bias in random walk

- Verification problems

Satellite-Aided Approaches

- Specialised (though open-source) software/s

- Professional equipment

- Spatial considerations focus mainly on catchment
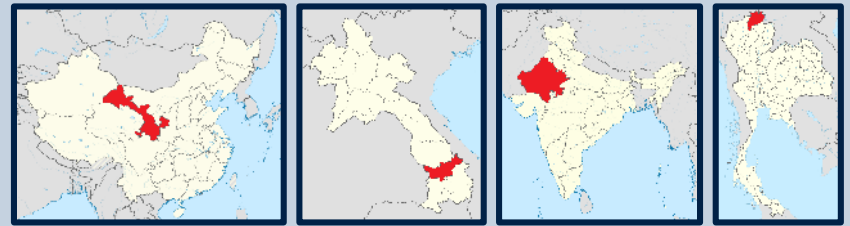
- Random walk issues remain

SETTING AND CHALLENGES

Setting and Challenges

Survey Data Collection

Survey data collected from 6,685 respondents across 97 PSUs (171 villages) in Gansu (16), Lao PDR (32), Rajasthan (16), and Thailand (33).

- **Eligibility: Ordinary village resident (>6m), 18 years and above**

- **4 steps in sample selection process**
  - Purposive selection of representative sub-districts in two districts
  - Random selection of 16/30 villages in these districts
  - Random selection of households in each village through interval sampling
  - Random selection of respondents for each household

- **60-minute questionnaires + PSU checklists**

- **Teams of 6 fieldworkers plus two supervisors**

Setting and Challenges

Survey Data Collection
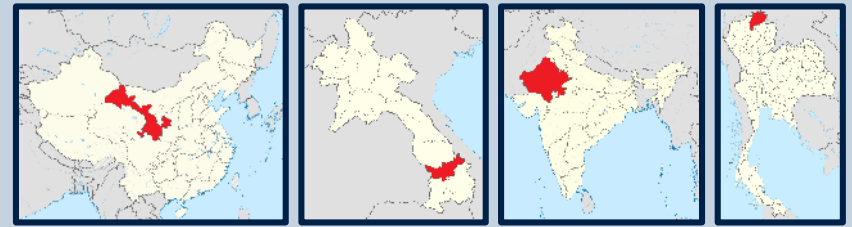
The rural surveys faced common administrative and resource constraints.

Administrative Challenges

- Absence of village population data (village name register from census)

- Mixed access to geographical information of villages

Resource challenges

- Varying labour and transportation costs

- Limited team time per village

PROCESS DESCRIPTION

Process Description

Village Selection (China)

Satellite maps were used to extract village locations for spatial stratification.

- **1,736 sites in eight sub-districts in register**
- **Locate villages in Google Maps using Chinese names**
- **Manually extract coordinates into Microsoft Excel**
- **Calculate distance to assigned and nearest township**
- **Drop or reclassify outliers**
- **1,553 villages in sampling frame, 16 plus 32 replacements drawn**

| Townships: 7		County Name	Latitude (google)	Longitude (google)	Distance to County	Distance to Town	Dist. to Nearest Township (km)
Village Name							
石青村委会	shí qīng cūn w	皋兰县	36.73834	103.7491	48.68	28.15	23.68
白坡村委会	bái pō cūn wěi	皋兰县	36.78682	103.8803	51.13	30.55	30.55
大横村委会	dà héng cūn wě	皋兰县	36.66784	103.8344	38.84	18.01	16.95
三和村委会	sān hé cūn wěi	皋兰县	36.55608	103.8584	26.29	5.65	5.65

Process Description

Village Selection (Laos, Thailand)

Satellite maps were used to extract village locations for spatial stratification.

- **No local administrative register but Country Files from US National Geospatial-Intelligence Agency**
- **Calculate distance of 3,100 villages to nearest district town**
- **Stratify and select sample**
- **Locate sample on satellite maps and resample if non-existent or unclear**
- **60 PSUs drawn, 54 PSUs resampled (16/30 in Lao Ngam / Ta Oy, Salavan)**
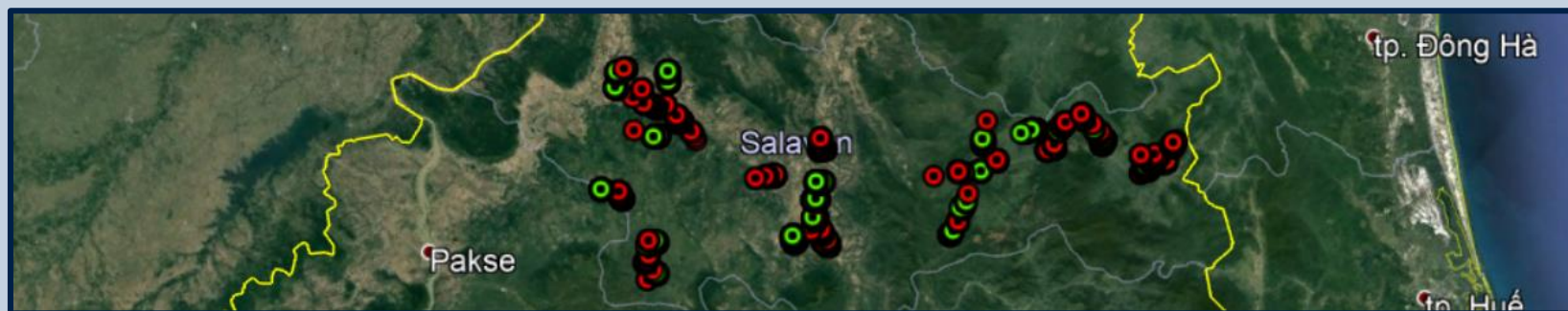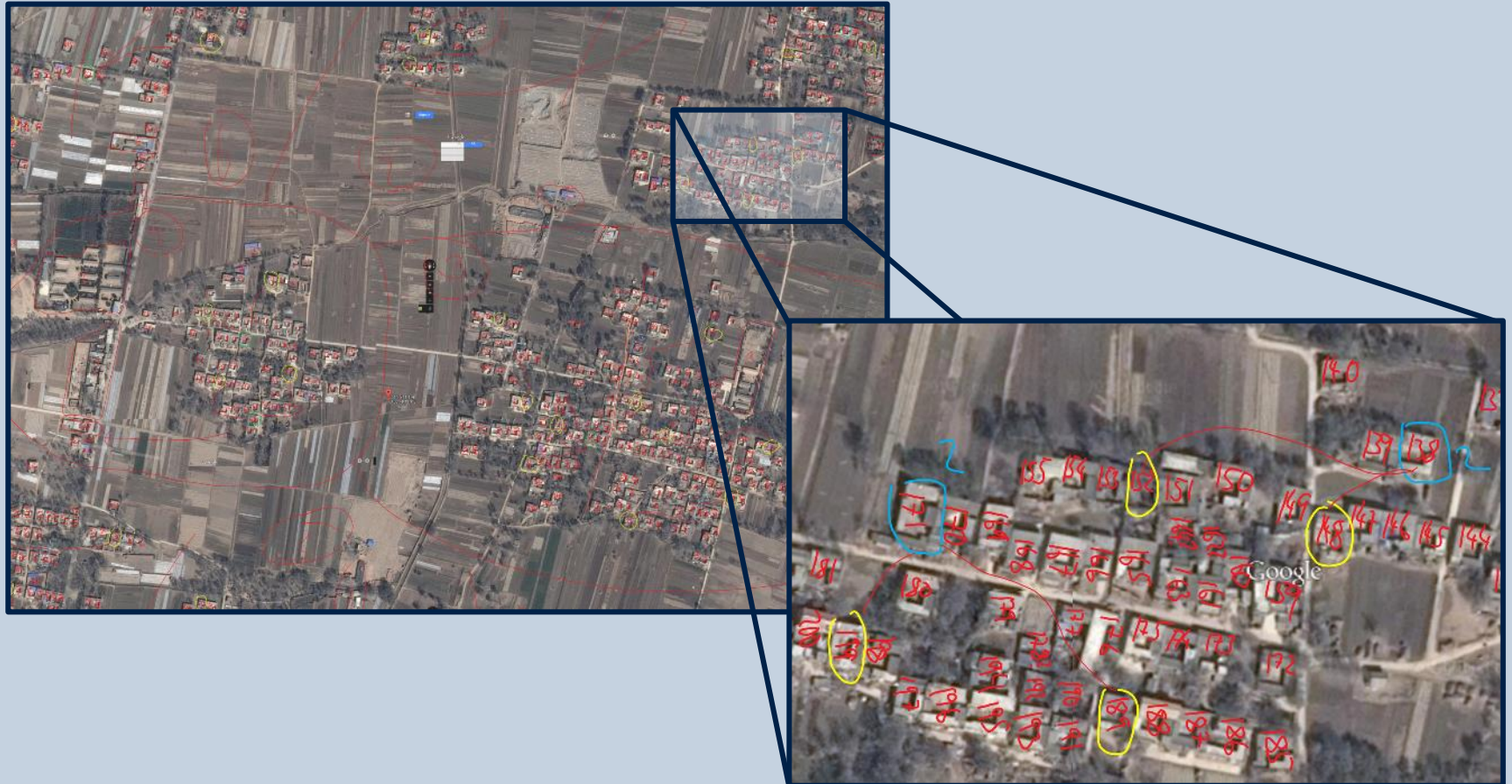
Process Description

Household Selection

High-resolution aerial maps of selected villages were extracted to build village-level sampling frames.

- **High-resolution village maps on Google Maps or Bing Maps**

- **Screen-cap highest-resolution maps (up to 1:670), 1km catchment**

- **Paste screenshots into Microsoft PowerPoint**

- **Assemble complete high-resolution area map (5-50 individual maps)**

- **Village maps contained total of 37,000 houses (up to 1,500 per village, 380 on average)**

Process Description

Household Selection

Process Description

Household Selection

Village-level sampling frames were built through complete enumeration of identifiable housing structures.

- **Homogeneous housing conditions, absence of apartment buildings**

- **Segment village maps to ensure spatial representativeness**

- **Number housing structures consecutively in PowerPoint**

- **Assign placeholders for unidentified structures**

- **Interval-sample (optionally plus replacement households) based on number of structures per segment**

Process Description

Implementation and Monitoring

Not only sample selection but also implementation logistics benefitted from digital aides.

- **Satellite maps facilitated upfront planning**
  - Village approach via chartered unchartered roads using Bing/Here Maps
  - Plan deployment/drop-off and pick-up
  - Assign investigators according to fitness in difficult terrain

- **Village maps to facilitate field investigator work**
  - Facilitate household location for fieldworkers
  - Field investigators to record GPS coordinates using dongles or tablets
  - Supervisors using geo-coordinates to verify correct households

DISCUSSION

Discussion

Benefits: Financial

Digital aides enabled low-cost sampling and streamlined logistics.

PhD Research Scenario (16 villages, 400 households, mountainous areas)

- **Savings ("realistic" scenario):**
  - Reduced travel, investigator work, insurance, researcher allowance
  - Saved 25% of main survey time, 80 labour days
  - **Total savings: £4,300**

- **Expenses**
  - Actual: Smartphone, printing, network charges
  - Optional: Touchscreen laptop, labour, GPS units, tablets for data collection
  - **Total expenditures: £170 to £2,000**

- **Bottom line: Savings of £2,300 to £4,100**

Discussion

Benefits: Methodological

ICT-aided sampling and implementation can also improve survey quality.

- **Complete listing, even for spatially marginalised households**

- **Spatial representativeness (superimposed grids)**

- **Less clustering, thereby improving effective sample size**

- **Stronger adherence to pre-selected households**

Discussion

Benefits: Methodological

In the absence of spatial autocorrelation, spatial stratification / segmentation is at least as accurate as simple random sampling.

Delmelle (2009: 186-189)
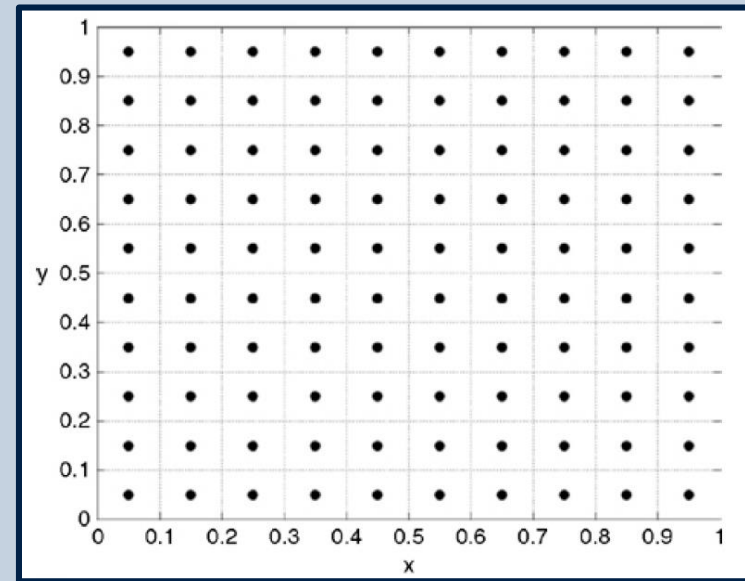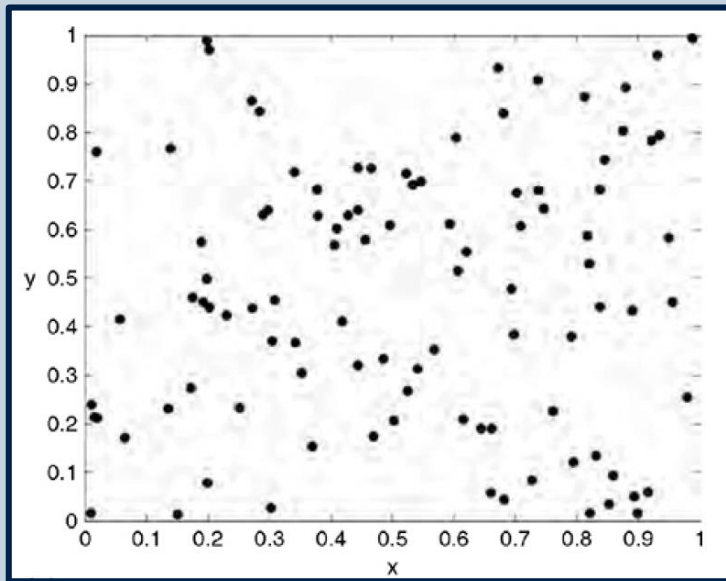
Discussion

Requirements

Researchers need to meet specific conditions to realise these benefits.

- **Locational requirements**
  - Local knowledge of field sites
  - Up-to-date, high resolution imagery
  - Identifiable, consistent, stable living conditions
  - Economic viability

- **Technical requirements**
  - Off-the-shelf ICT
  - 2G connectivity (not needed for GPS signals)
  - Optional: Lowest-cost GPS units, compasses, tablets

- **Logistical requirements**
  - Training on map reading
  - Detailed team instructions prior to village visit
  - Information from local leaders

Discussion

Challenges

Reliance on technical aides can have methodological implications.

- **Village selection**
    - Use of official but incomplete village registers can curtail sampling frame
    - Insensitivity of population size can produce bias towards smaller villages
    - Sensitivity towards population size can explode survey costs
    - Villages outside of selected district need separate permission for access

    - **Remedies:** Visual inspection of maps, expand catchment areas, ex-post weighting, replacement village list (readily sampled)

- **Household selection**
    - Hit rates between 50% and 94%, average of 79% are residential buildings
    - Abandoned and shared housing units can influence sampling interval
    - Manual household listing can create trust that is forgone in ICT-aided sampling (total of 61 refusals or up to 4 per village)

    - **Remedies:** on-the-spot updates of village maps, revisits, drone reconnaissance

CONCLUSION

Approaches to Survey Sampling

Satellite-Aided Approaches

- **Useful approach in contexts that have**
    - Lack of administrative data for sampling frame
    - High survey costs
    - Homogeneous and identifiable dwelling units
    - Quality satellite maps (Google, Bing/Here, Apple, …?)

- **Transparent, efficient, low-cost alternative to random walk and household listing**

- **Saved between £4,000 and £39,000, making surveys feasible**

- **No specialised equipment / skills required**

- **Suitable for student researchers and resource-constrained studies**

More detail here (open access): Haenssgen, M. J. (2015). Satellite-aided survey sampling and implementation in low- and middle-income contexts: a low-cost/low-tech alternative. *Emerging Themes in Epidemiology, 12*(20). doi: 10.1186/s12982-015-0041-8

**THANK YOU.
QUESTIONS?**

marco.haenssgen@ndm.ox.ac.uk

Twitter: @HaenssgenJ