

# Antibiotics and Activity Spaces: An Exploratory Study of Behaviour, Marginalisation, and Knowledge Diffusion

## Data Note

Marco J Haenssger, 13 March 2019

These three data sets comprise each a provincial-level representative rural survey of adults in Chiang Rai and Salavan, and a two-round census survey with a three-month interval in five villages across the two sites (3 in Chiang Rai, 2 in Salavan). The surveys were implemented by 10-member survey teams in each country between November 2017 and April 2018.

**1) Provincial-level representative survey:** Three-stage stratified cluster random sampling design. The first stage involved the random selection of 30 primary sampling units (PSUs) across five purposively selected districts in each site, stratified by their distance to the nearest urban centre (using data from the National Geospatial-Intelligence Agency, 2017). The second stage enumerated all residential buildings within the selected villages using satellite imagery from Google Maps and Bing Maps, of which we sampled 5% of the buildings (but at least 30 houses) in a stratified interval sampling approach to ensure spatial representativeness. During the survey implementation, the third sampling stage involved selecting randomly one respondent for every five adults in each chosen house. The resulting data are representative for the general adult population of rural Chiang Rai and Salavan (522,000 in Chiang Rai and 190,000 in Salavan as per census data).

**2) Two-round village census:** Community-level social network census surveys in five purposively selected villages across the two field sites (3 in Chiang Rai, 2 in Salavan). The villages were selected in consultation with local stakeholders; guiding criteria were (1) village size and structure, (2) remoteness and road accessibility, (3) economic status as approximated by village-level infrastructure and facilities, (4) ethnic composition and (5) number and location of health facilities within a 2 km radius. The villages had between 300 and 1,500 residents. Within the selected communities, all households were approached, their adult members enumerated and invited to participate.

The village and household selection used a satellite-aided sampling approach (Haenssger, 2015). We used a list of 3,100 geo-coded villages across the two field sites, which we stratified according to the district-level median distance to the nearest urban area (which could be within or outside of the same district). We randomly sampled 60 PSUs from the stratified list (30 each above and below the district median distance to the nearest town). However, the data from the US National Geospatial Intelligence Agency was not always accurate in locating actual villages. In the case that a set of coordinates did not actually corresponded to an existing village, we resampled that PSUs randomly until a village was located. The resampling was necessary especially in the Salavan districts of Lao Ngam and Ta Oy (which contained many dispersed settlements), where we had to resample 16 and 30 times, respectively (or 85% of the total 54 occasions in which resampling was necessary).

Ultimately, we selected 33 PSUs in Chiang Rai (incl. 3 census villages) and 32 (incl. 2 census villages), which represented 72 administrative villages in Chiang Rai and 69 in Salavan. We used Bing and Google satellite maps as well as Google Street View in Thailand to identify and enumerate housing structures, which served as basis to calculate village size and to select the random interval sample for each village. We manually enumerated more than 30,000 housing structures using this method. The accuracy of identifying dwelling units rather than other types of houses ranged from 50% to 94% per village (79% on average).

Following the introduction of the field team to each village through public announcements by the village heads, the team members – occasionally accompanied by local officials for monitoring – approached the sampled houses and invited a randomly chosen household member to participate in

the interview (the randomisation was operated through the survey software SurveyCTO, through which we administered our questionnaires; the census surveys enumerated all household members and invited all adults of at least 18 years to participate in the survey). The field team performed up to two visits to each sampled house, and they were encouraged to make appointments or locate the selected household member in or around the village if they were unavailable at the first visit (our household definition was based on a shared kitchen and previous residence of at least six months in the survey village). In case a sampled house was not inhabited or unavailable, the nearest neighbour was chosen to preserve the geographic dispersion of the interval sample. The recruitment and data collection process was monitored in the field through the survey supervisors and a research officer (who conducted spot checks), and remotely via the SurveyCTO monitoring tools that enable visualisation of the interview locations (thereby minimising risks of data fabrication) and speedy queries and feedback on submitted data. 17.6% of all interview sessions were attended by a survey supervisor for monitoring purposes. Where the quality of a survey interview was not satisfactory or where the digital data entry was corrupted, field investigators and survey supervisors were tasked to correct the errors and revisit the respondents (which was the case in less than 20 instances).

Our survey instrument was a face-to-face questionnaire, administered on tablets running SurveyCTO (Dobility Inc., 2017). The field teams were recruited locally and comprised six to eight field investigators plus two survey supervisors. The team received five days of full-time classroom and field training (plus an additional five days for supervisors) in order to sample houses and administer the questionnaires in the local language (Thai/Lao). Issues with minority languages were rare (228 instances) and resolved by recruiting local translators within the villages. The duration of the survey interviews ranged from 15 to 60 minutes depending on the type of the survey and the detail of information that the participants volunteered. For instance, the first round of the census survey involved constructing social network relationships within the village, which tended to consume more time than the provincial-level representative survey, where this module was omitted. Of all respondents, 46.9% volunteered information about at least one recent illness episode.

The provincial-level representative survey involved 2,141 responses in total; the census surveys gathered a total of 3,744 responses (from more than 2,000 villagers over two survey rounds). The census surveys in Chiang Rai managed to re-interview 87.5% (548) of the first-round respondents and 89.4% (1,130) in Salavan, whereby seasonal outmigration limited the opportunities to re-visit first-round respondents. We chose an oral consent procedure to not alienate and discriminate against illiterate or unregistered participants, and we offered the option to opt out of data sharing. Of the 5,885 responses in total, we did not receive consent for data sharing in 15 instances. These responses have therefore been removed from the public data set.

The public data sets contain information from all survey types on the individual level, on the illness level for those cases where respondents reported a recent illness that they or a child under their supervision encountered, and on the “step level” (containing for instance information on medicine use episodes at different stages of an acute illness or accident). The public data sets have been anonymised, meaning that we removed geographic identifiers, names of respondents and social network contacts, and interview notes that would permit deductive disclosure of our informants’ identities. Free text responses in Thai and Lao were translated into English by the project team, and the data sets include both original and translated text for full disclosure.

## References

- Dobility Inc. (2017). *SurveyCTO*. Cambridge, MA: Dobility Inc.
- Haenssger, 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
- National Geospatial-Intelligence Agency. (2017). *Country Files (GNS)*. Retrieved 17 July 2017, from <http://geonames.nga.mil/gns/html/namefiles.html>