

# DOCUMENTATION GUIDE #1: SAMPLING PROTOCOL



## TLT Sampling Protocol

Tsogolo la Thanzi (TLT) placed considerable emphasis on drawing a genuinely representative and random sample of respondents. In order to achieve this goal, we conducted a complete household census of all households in census enumeration areas (EAs) where at least half of the EA was within 7 kilometers of the center of Balaka town (see Figure 1). Based on Malawi's 2008 Census, this was approximated to be 64,000 people residing in 15,000 households in 54 EAs. Our aim was to capture all residents of the catchment area in May 2009, using the standard of "usual residence" (i.e., those residents who usually reside in a household). Accordingly, we did not list students whose families live in Balaka but who board outside of the district. Students who board within the catchment area were listed at their school, not within their homes. We took great care to include non-traditional populations. We listed the largest boarding school in the area and listed regular residents of rest houses (including bar owners and bar workers, who are often sex workers as well). The only populations known to be absent from the sampling frame were female boarders at the local Catholic all girls high school (~250) and a population of male boarders (~250) at a local Catholic all boys school. The administrators at the girls' school did not allow us to list their students but did inform the girls of the study and told them they could independently register to participate at the Tsogolo la Thanzi center.

The success of the household listing was contingent upon approval from local authorities (both traditional and governmental). Prior to beginning the listing, TLT management met with the traditional authority for the entire area, local chiefs, the district commissioner, and the district health officer. At these meetings, we explained the project's purpose and protocol and dealt with questions and concerns that the authorities had. Additionally, TLT personnel held village meetings, sanctioned by the local chief, where they explained the household listing and recruitment process to interested residents. Without exception, problems that arose during listing and recruitment (e.g., questions about the legitimacy of the project and the true purpose of the study) occurred in particular villages where the meeting had been postponed/missed because of a local funeral.

Jenny Trinitapoli and Sara Yeatman designed the household listing questionnaire jointly, with input from Susan Watkins and Abdul Chilungo. The relatively short questionnaire was completed by the household head or other adult resident; household informants provided basic information on every individual who was usually living in that household and answered a short set of questions about household goods and infrastructure. The interviewer also assessed household roofing material.

Figure 2 presents the age and sex structure of the Balaka catchment area. At the time of listing, the area contained 6696 eligible women (born between 1984-1994, inclusive) and 6180 men (born between 1984-1994, inclusive<sup>1</sup>). All household listings were entered into a simple database. Eligible individuals were treated as independent observations and created the sampling frame. 1500 women were randomly selected

---

<sup>1</sup> The administrators of the boys' section of the government boarding high school in the area did not have access to birth year data. All male high school students were added to the sample frame and their birth year was checked against the eligibility requirements during recruitment.

for inclusion in the sample using the random number generator in Stata. The only people excluded at the time of the random draw were members of the immediate TLT staff and their spouses.<sup>2</sup> At selection, an additional 200 women were marked as reserve respondents to compensate for ineligible women (i.e., due to misreporting of age during the household listing) or refusals. A similar protocol was followed for the 600 male respondents drawn from the general population (as opposed to our respondent-driven partner sample). The sampling ratio is approximately a 1 in 4 for women and 1 in 10 for the random sample of men.

After being drawn into the sample, respondents were sorted by “zone” (i.e., location within the sampling area) and randomly assigned to a primary interviewer, matched by gender. Each TLT interviewer was responsible for recruiting and interviewing a roster of approximately 100 respondents. Each interviewer recruited his/her respondents in their homes, obtained parental consent for minors, and invited respondents to the TLT Research Center to complete their baseline interview. Interviewers set appointments with each respondent and gave respondents calendars to remind them of their baseline interview date. At the time of recruitment, respondents were considered ineligible if they were outside of the age range, were permanent migrants, or were deceased. Respondents were considered non-responders if they refused to participate in the study, were not found after three attempts to locate them, did not show up to their scheduled interview or to the second attempt to reschedule (conservatively considered passive refusal), or were deemed unable to participate because of language (i.e., not speaking one of the three dominant local languages) or severe mental disability. The female response rate was 96% and male response rate was 94% (see Table 1).

---

<sup>2</sup> Consistent with our sampling ratio, we drew 4 out of 16 eligible female staff members into our sample and 3 of 12 age-eligible male staff members. These individuals were dropped from the sample and replaced with people from our lists of reserve respondents.

### **Notes to Analysts on Using the Household Listing Dataset Responsibly**

The household listing dataset is in long format. Analysts interested examining these data by household (the original unit), should reshape the dataset on `hhid`.

Due to irreconcilable doubts about household of origin, there are three respondents in the TLT study (Wave 1) for whom we did not make links to their place in the household listing (`respids`: 110829, 129712, 5220306).

First Drafted By: Jenny Trinitapoli and Sara Yeatman, June 10, 2009

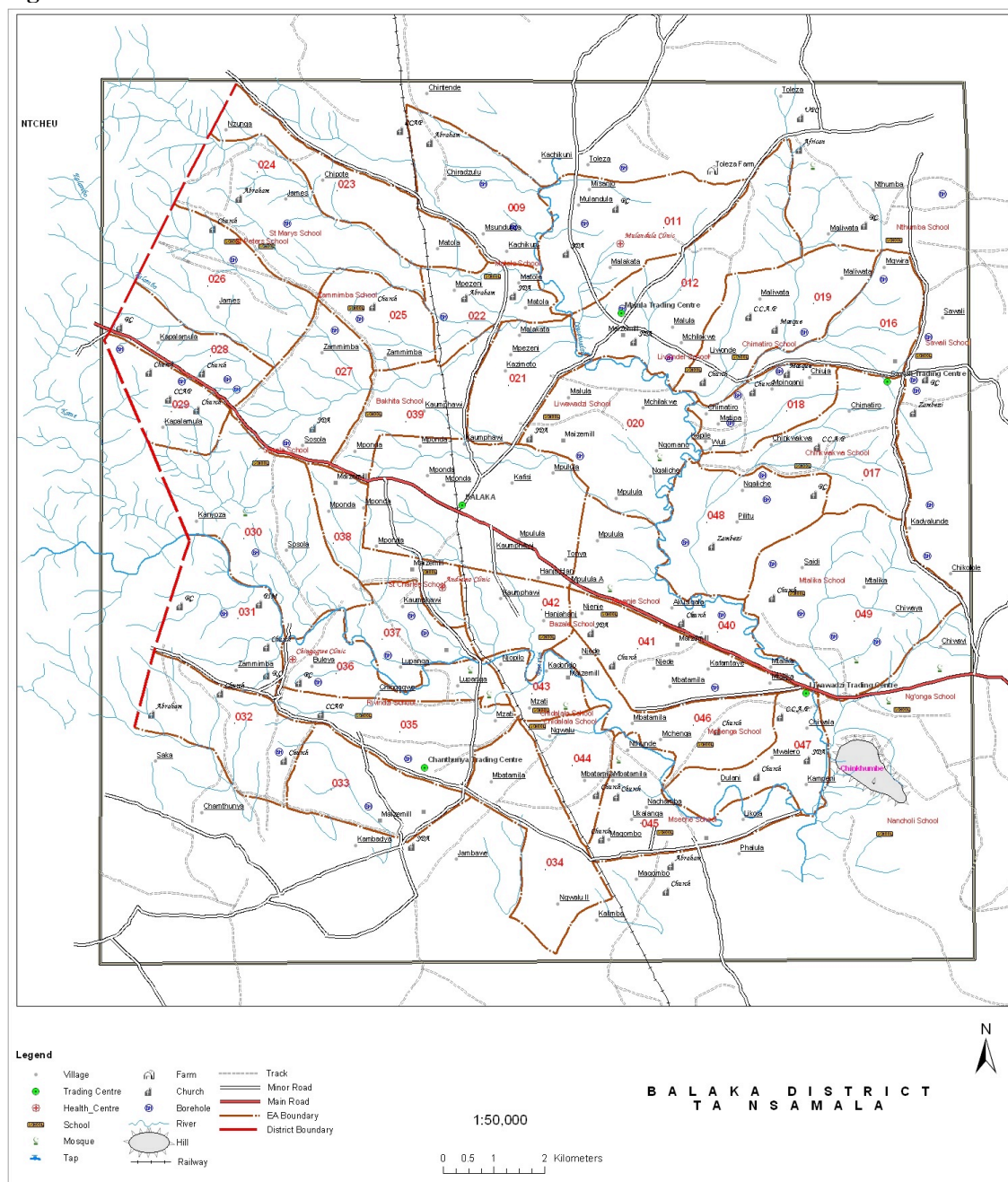
Revised & Clarified Version: Nell Compennolle, December 3, 2018

## Tables & Figures

**Table 1. TLT baseline response rate, May-August 2009**

	Females	Males
Completed	1505 96.4%	574 93.5%
Not found	17 1.1%	21 3.4%
Refused	30 1.9%	11 1.8%
Language/mental problems	10 0.6%	8 1.3%

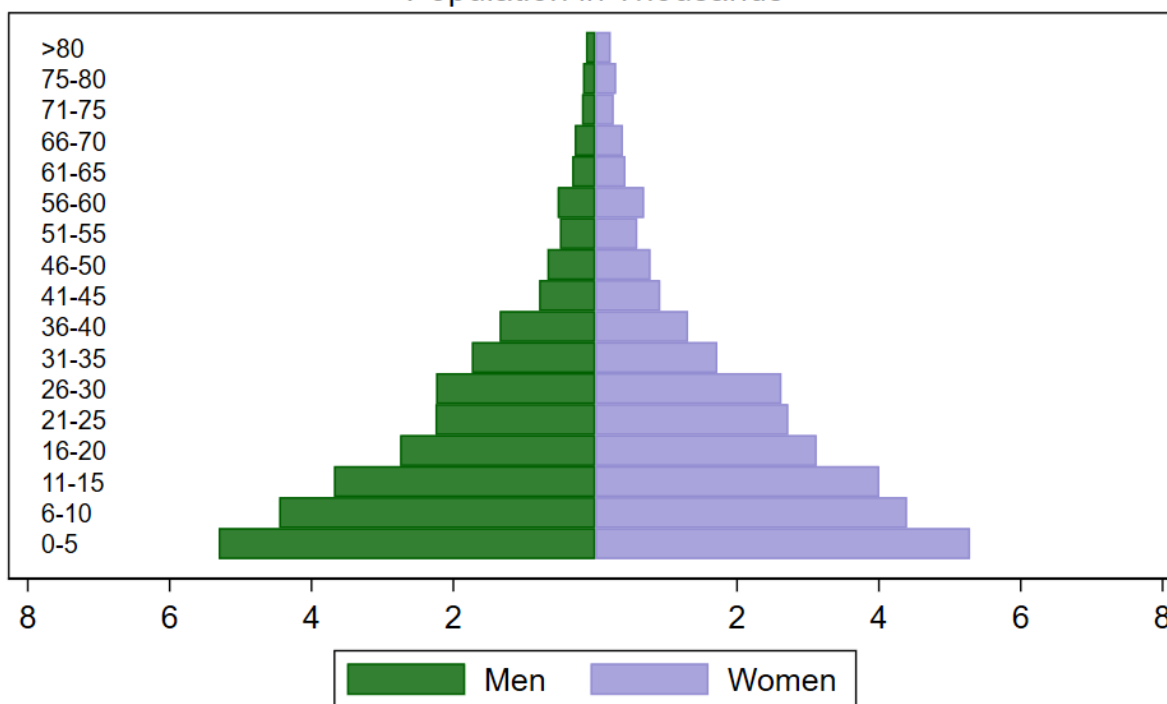
Figure 1 TLT Catchment area



**Figure 2 Age and Sex Structure of the Balaka Catchment Area**

## Population of Men and Women in Balaka by Age, 2009

Population in Thousands



Source: Tsogolo la Thanzi 2009