

Whole Village Project Methods for Secondary Data Analysis Users

Table of Contents

1. Purpose of this document.....	2
2. Background on the Whole Village Project	2
3. Data collection methods and survey tools.....	3
3.1 Village level sampling.....	3
3.2. Household Sampling	4
3.3 General procedures in survey data collection and basic cleaning.....	4
3.4. Protocols for Quantitative Survey Participant Selection	5
3.4.1 Protocol for Household survey	5
3.4.2 Protocol for Nutrition, Food Security and Jatropha survey	5
3.4.3 Protocol for Under-five Nutrition Survey.....	5
3.4.4 Protocol for HIV/AIDS Knowledge, Attitude and Practice (KAP).....	6
3.5 Qualitative data.....	6
3.5.1 Focus Groups.....	6
3.5.2 Key Informant Interviews.....	7
3.5.3 Data recording	7
4. Data Cleaning and Merging.....	7
5. Acknowledgement of Data Source by Future Users	8
Appendix – A Table of Villages by District and Region.....	10
Appendix B – Sources of Survey Questions	12

1. Purpose of this document

The purpose of this document is to provide background information on data collection and methods for researchers conducting secondary data analyses. The data collection methods have remained the same throughout the project. However, there have been some changes in the survey tools over the life of the project. All survey tools and iterations, a Masters' Variable documentation on the variable name changes for each survey, and data dictionaries can be found on the WVP website

<http://wholevillage.umn.edu/> on the Internal Resources link. This is a password protected area. Email or call individuals on the contact information page for the current user name and password.

2. Background on the Whole Village Project

The Whole Village Project (WVP) began collecting and analyzing comprehensive data at village level in Tanzania in 2009. The current data set includes quantitative and qualitative data from 56 villages collected between 2009 and 2011. The WVP is a collaborative project between Savannas Forever Tanzania (SFTZ), a Tanzanian NGO, and the University of Minnesota, USA. The **vision** of the WVP is to work with people in rural Tanzanian villages to acquire and use knowledge for improving long-term health and well-being while sustaining natural resources. To achieve this goal, quantitative and qualitative data are systematically collected in villages across northern Tanzania by the Savannas Forever team in partnership with staff from the National Institutes of Medical Research (NIMR) and the Tanzanian Wildlife Research Institute (TAWIRI). The data are sent to the University of Minnesota for data management, storage, cleaning and analysis and then returned to Tanzania. The SFTZ team returns to each village to present the data to villagers for their own use and decision-making. WVP intends to return to each village every two to three years in order to assess the sustainability of development projects over time and identify best practices.

The Whole Village Project's survey tools and methodology have been reviewed and approved by multiple Tanzanian research authorities (COSTECH, NIMR and TAWIRI) and the University of Minnesota Institutional Review Board for the ethical conduct of human subjects' research. Further, permissions are sought by the respective regional, district and village leadership before beginning data collection.

Many of the survey tools and questions were first piloted by SFTZ between 2006 and 2008 and formed the core household survey at the initiation of the WVP. During 2009 to 2011, SFTZ received funding from PEPFAR/Tanzania to conduct the core survey and an HIV/AIDS Knowledge, Attitude and Practice (KAP) survey in 56 villages. In addition, SFTZ received matching funds from Partners for Development (PFD) to conduct specific baseline data collection on jatropha production, nutrition and food security information in 20 of the 56 villages. Other matching funds were received from Global Service Corps (GSC) for additional questions on kitchen gardens and agricultural techniques; African Wildlife Foundation, Frankfurt Zoological Society and Friedkin Conservation Fund on wildlife conflicts for villages in proximity to Wildlife Management Areas (WMAs). For each different funder, additional modules were added or dropped from the core survey. After conducting data collection in the first 14 villages in 2009, a number of changes were made to the core survey in March 2010 primarily to questions related to income. Further, some questions were and dropped or added to improve data quality. A few questions in each

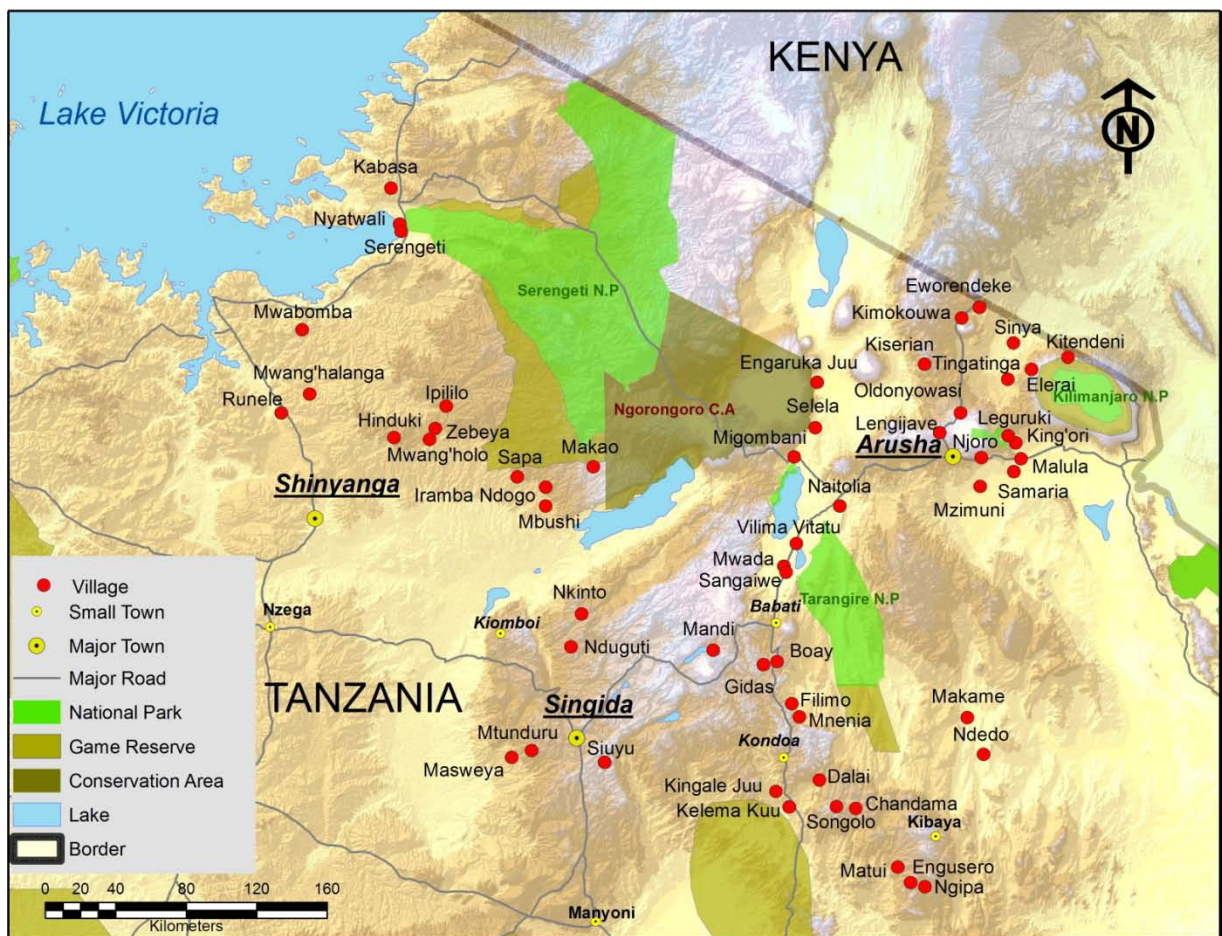
survey were added or dropped in the following months. As a result, there are multiple iterations of the surveys used. A summary of all variable names and any changes with each survey iteration have been recorded and can be referenced along with the data dictionaries.

3. Data collection methods and survey tools

3.1 Village level sampling

Village selection is based on the funding agency priorities and permission of government leaders. After permissions are received the Savannas Forever Tanzania (SFTZ) staff arranges dates for data collection with district officials and village leaders.

A map of the 56 villages included in this sample is shown below.



Appendix 1. Provides a table listing each village and the Tanzanian Administrative district and region where each is located.

3.2. Household Sampling

Village leaders provide a roster list of heads of households and the research team uses a computer generated randomization program to select 60-75 households from this list. A Tanzanian survey team of 6-7 personnel work in each village for 5-6 days. The team begins with a sensitization session with leaders and community members to introduce the project and staff.

3.3 General procedures in survey data collection and basic cleaning

All quantitative data were collected through face to face interviews. The surveys were programmed into a personal digital assistant (PDA). Each interviewer read aloud the survey from the PDA and coded responses using a drop menu. A team supervisor worked with a group of 4-6 interviewers. At the end of each day, the team met and reviewed any problems with interviews or specific questions on the survey. The team supervisor randomly selected 10% of households for quality control checks. She or he would return to the household and select a sub-set of survey questions to repeat them and compare the results with the original interview data.

At the end of data collection in each village, the team supervisor downloaded the PDA data into an excel spreadsheet. Preliminary data cleaning checks were conducted in the field to check for missing data or errors with data entry. Once the data checks were complete, the village data files were sent to the University of Minnesota (UMN) WVP research team via encrypted email as excel files. The UMN research assistant (RA) receiving the file logged a record of each file and saved an original in an archive file and a second original with all identifiers (names and GPS units) removed. A separate copy was used for data cleaning and analysis.

For the first 14 villages, data were first cleaned in excel and then uploaded into SPSS for analysis. After survey changes were made in March 2010, the research team decided to transition all data cleaning and analysis into Stata 11.0. Data available on our internal resources web page allow users to either download data as Ascii files or as Stata files for transfer into your preferred statistical package. These data have already been cleaned. Four quantitative surveys are used across Whole Village Project villages:

1. Household or SES survey
2. Nutrition, Food Security and Jatropha
3. Children Under-five
4. HIV/AIDS Knowledge, Attitude and Practice (KAP) survey

The first survey is referred to as the Household Survey or sometimes as "SES Survey". These data arrive in two files: one is the Household Demographic data or roster; and the second is the remainder of the Household survey referred to as "SES". The Household Demographic file contains information on the Household Head (HH), all people living in the household regularly, their relationship to the HH, age, sex and level of education completed. There is a unique Household Number for each household within a village; however, **only in villages 15 on is there a unique Person Number Identifier** for each member of the household. This becomes relevant if you want to merge data from the different surveys and compare a person in say the KAP survey to data from their SES data or the Household Demographic

data. For ALL villages, data for individuals in the Children Under-5 survey and the HIV/AIDS KAP survey can be linked to Household Data through the Household Number. But only in villages 15+ can one link individual person data (such as occupation and level of education).

The questions in the survey tools are largely taken from previously validated publicly available sources and are described in further detail in Appendix B.

The protocol for each survey tool is described in each section below.

3.4. Protocols for Quantitative Survey Participant Selection

All interviews are conducted in Kiswahili whenever possible. If respondents are not fluent in Kiswahili, a bi-lingual villager is identified by the leadership to translate from the local language to Kiswahili. At the end of the household survey there is a section for interviewers to complete indicating if the survey was completed, the length of time, if there were problems and if there was a translator.

3.4.1 Protocol for Household survey

As noted at the beginning of section 2, every household is randomly selected to be interviewed based on a randomization of all households across sub-villages within a given village. Village key informants accompany the interviewer and driver to show them where each household is and introduce the interviewer to household members. Households may be clustered together or several kilometers apart. The villager key informant does not remain for the interview.

The participant selection protocol for the Household Survey is to interview the Household Head. In most households, this is the male head of household. If he is away from the house and not available, and another adult, usually the HH's wife is home and available, she may be interviewed. If there are sections of the Household Survey she is unable to answer, the interviewer is to return to the household when the HH is available.

Female-headed households usually occur under three primary conditions: (1) the female-head has never been married or is widowed or divorced; (2) the woman is part of a polygamous household and lives under a separate roof/house from her husband and serves as "head" of that household; (3) the woman is married but her husband lives elsewhere most of the year (usually working in another part of the country) and so she plays the primary role as household head.

Each household receives a half kilo of sugar for their time as compensation.

3.4.2 Protocol for Nutrition, Food Security and Jatropha survey

After the Household Survey is completed, the interviewer moves on to the Food Security and Nutrition Survey. As most of the questions on this survey are related to diet, kitchen gardens and firewood collection, the female "head of household" is the preferred participant to answer these questions as typically; she would perform many of these tasks and answer the questions most accurately.

3.4.3 Protocol for Under-five Nutrition Survey

If the selected household has any children under the age of five, an additional survey is conducted to inquire about **each** child's diet, breastfeeding practices, immunizations and their height and weight. The

interview protocol is to ask the female head of household (usually the mother) or primary caretaker for the child to participate.

Given time constraints and stress on a household, no more than five Under-five surveys are to be conducted per household. If there are more than five children in the household, then the name of each child is given to the interviewer and she randomly selects up to five children for measuring and for the survey.

All children are weighed on a hanging scale. For height and length measurements and standardized metrics a standing board is used for children who can stand. For infants, the board is laid flat and the infant's length is measured on the board.

3.4.4 Protocol for HIV/AIDS Knowledge, Attitude and Practice (KAP)

The HIV/AIDS KAP survey is conducted with up to 4 adults (15 years or older) within the household. Interviewers are encouraged to non-randomly identify up to two males and two females per household. All interviews are conducted in a private space where no one else may listen. Given that more women are at home during the day when the surveys are conducted, there are more women in the data set than men.

3.5 Qualitative data

In order to obtain more contextual data about each village, a number of focus group and key informant interview tools are used. A general description of the protocol for participant selection is described here.

3.5.1 Focus Groups

As noted earlier, the Savannas Forever Tanzania team leaders' first meet with Village Executive Officers before the team arrives in order to assure the team has the permission/clearance from the village leaders and secure the best dates for the village visit. In addition, village leaders are asked to invite villagers to participate in the multiple focus group sessions. These include the men's and women's focus group and the Institutional Analysis. As staffing allows, most focus groups are conducted by a team of two staff with one facilitator and one note taker. Village heads are encouraged to invite a mix of villagers from across sub-villages to participate. For the men's and women's focus groups, a mix of ages is encouraged. Most of these focus groups have 8-10 participants and are conducted as a single sex group.

The Institutional Analysis focus group includes a mix of men and women of 8-10 participants. Usually they are a mix of ages and professions although occasionally they are heavily represented by village leaders and occasionally only men.

For the Agricultural and Livestock Focus Group, village leaders are asked to invite some of the most respected or well known farmers or livestock holders across the sub-villages. The participants are asked to give general information about agricultural and livestock issues across the village so having the most knowledgeable participants is of value.

The Village Head focus group provides information on village institutions and services. The participants include the Village Executive Officer and the sub-village chairmen. Most participants are male.

The Village Resource Focus Group is usually smaller and includes 4 to 6 individuals. Both men and women are able to participate depending on their role. If there is a Village Game Scout, he or she is invited to attend along with farmers and/or village leaders who are knowledgeable about village natural resources and institutional resources.

3.5.2 Key Informant Interviews

There are two key informant interview guides: one for school headmasters and a second for a clinic officer. For headmasters, the protocol is to conduct interviews with the headmaster of every school in the village, both private and public (government) and for primary and secondary schools. One SFTZ researcher conducts the interview and takes all the notes. The researcher is encouraged to try and conduct a physical tour of each school in order to add notes about physical conditions.

If there is a clinic or dispensary in the village, a SFTZ researcher asks to conduct an interview with the most senior health personnel available, usually a medical officer or a nurse.

3.5.3 Data recording

All qualitative focus group and interview notes are recorded by hand in Kiswahili on a sheet of paper that follows the focus group or key informant discussion guide. Within two weeks of the interview, the note taker enters the data in an English excel spreadsheet with the answers recorded on separate tabs. The data entry person does a translation from Kiswahili to English during the data entry phase. A second staff person reviews the data entry and the notes to assure for accuracy in translation and completing the form correctly. As there is no voice recording of the focus groups, we have no direct quotes from participants, just paraphrases from the recorder/translator.

4. Data Cleaning and Merging

The data cleaning principles have been the same throughout the data collection process. As noted earlier though, in the first 14 villages the data were cleaned in Excel and then entered into SPSS for data analysis. Since March 2010, all data have been cleaned in Stata 11.0. Data cleaning has focused on assuring that all written variables such as village names are consistently spelled; removing duplicate files; matching household and person number data across the different surveys (e.g., the SES household survey with the Under-five or HIV KAP surveys); and assuring there is just one head of household per household.

As noted earlier, the data from villages 1-14 and 15-56 are now fully merged. The individual data files for each survey (HIV KAP, Food Security, etc) however, are not merged. So as a future user, you will need to merge the different survey data sets appropriate to your research question. Most commonly, users merge the data sets using Village Codes and Household Number codes.

5. Acknowledgement of Data Source by Future Users

The Whole Village Project would like these data to be used by Tanzanian and international project managers and researchers in order to better inform development strategies in rural Tanzanian communities.

Researchers and program managers are asked to appropriately acknowledge the source of these data in the following ways:

- Researchers acknowledge the source of the data in any oral presentations or written publications with the following statement, “We thank the Whole Village Project (Savannas Forever Tanzania, the Tanzania National Institute of Medical Research and the University of Minnesota) for the use of these data.”
- Researchers provide a copy of the research findings to WVP/University of Minnesota, SFTZ and the National Institute of Medical Research including conference presentations, unpublished dissertations or master’s theses, unpublished reports and any publications.
- Researchers agree that they do not have exclusive access to the data with respect to any specific research question.
- Researchers agree to use the following citation,

Whole Village Project. *Whole Village Survey, Tanzania* [Computer file]. Savannas Forever Tanzania, Tanzania National Institute of Medical Research and the University of Minnesota [producers]. Minneapolis, MN: University of Minnesota [distributor], 2010.

Unless changed on our website, copies of reports, papers or oral presentations should be sent to **each organization and contact person listed below**:

Dr. S. Mfinanga, MD, PhD
Centre Director and Principal Research Scientist
National Institute of Medical Research
Muhimbili Medical Research Centre
PO Box 3436
Dar es Salaam, Tanzania

Email: gsmfinanga@yahoo.com or gsmfinanga@nimr.or.tz

Ms. Susan James, Executive Director
Savannas Forever Tanzania
PO Box 878
Arusha, Tanzania

Email: susan.james@sftz.org
<http://www.facebook.com/savannasforever.tz>

Dr. Kari Hartwig, DrPH, WVP Program Director
University of Minnesota
Global Programs & Strategy Alliance and Minnesota Population Center
50 Willey Hall
225 19th Avenue South
Minneapolis, MN 55455

Email: khartwig@umn.edu

Appendix – A Table of Villages by District and Region

Table 1. Overview of villages by district and region

Region	District	Villages
Arusha	Meru	King'ori
		Leguruki
		Samaria
		Malula
		Njoro
	Arusha rural	Oldonyowas
		Mzimuni
		Lengijave
		Kimokouwa
	Longido	Eworendeke
		Tingatinga
		Elerai
		Kiserian
		Sinya
		Kitendeni
		Engaruka Juu
		Migombani
		Selela
		Naitolia
Manyara	Babati	Boay
		Gidas
		Mandi
		Mwada
		Sangaiwe
		Vilima Vitatu
	Kiteto	Ngipa
		Matui
		Engusero
		Makame
		Ndedo
Mwanza	Kwimba	Mwabomba
		Mwang'halanga
		Runele
Shinyanga	Maswa	Hinduki
		Ipililo
		Mwang'holo
		Zebeya
	Meatu	Mbushi

Region	District	Villages
		Iramba Ndogo
		Sapa
		Makao
Singida	Iramba	Nduguti
		Nkinto
	Singida rural	Masweya
		Mtunduru
		Siuyu
Dodoma	Kondoa	Kelema Kuu
		Filimo
		Mnenia
		Kingale Juu
		Dalai
		Songolo
		Chandama
Mara	Bunda	Kabasa
		Nyatwali
		Serengeti

Appendix B – Sources of Survey Questions

The majority of survey questions come from previously validated publicly available survey tools. For each survey tool, we briefly review the sources of questions by content area.

Household Socio-economic Survey

Socio-economic, occupation, education, agriculture, house structure, assets, water sources, cooking stoves, mosquito nets, household shocks

The two primary sources include

(1) the World Bank's Living Standards Measurement Survey
<http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html> and the

(2) Demographic Health Survey <http://measuredhs.com/>

Food Security, Nutrition, Jatropha and Kitchen Gardens

Nutrition

Demographic Health Survey

Food Security

Deitchler, M., Ballard, T., Swindale, A., and J. Coates. 2010. *Validation of a Measure of Household Hunger for Cross-Cultural Use*. Washington, D.C.: Food and Nutrition Technical Assistance II Project (FANTA-2), Academy for Educational Development, May 2010.

Coates, J., Swindale, A., and P. Bilinsky. 2007. *Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide (v.3)*. Washington, D.C.: Food and Nutrition Technical Assistance Project, Academy for Educational Development, August 2007.

Jatropha production and Kitchen Gardens

Questions developed by SFTZ and University of Minnesota staff and faculty in collaboration with Partners for Development and Global Service Corps staff.

Under-5 Nutrition

Demographic Health Survey

Additional questions developed by Dr. Esther Ndagaya, National Institutes of Medical Research, Tanzania and Dr. Monique Borgerhoff Mulder, University of California-Davis.

HIV/AIDS Knowledge, Attitude and Practice Survey

Demographic Health Survey