NATIONAL BUREAU OF STATISTICS

Basic Information Document

National Panel Survey (NPS 2008-2009)

2008-2009

ACRONYMS

BMGF	-	Bill & Melinda Gates Foundation
DECRG	-	Development Economics Research Group
DFID	-	United Kingdom Department for International Development
EA	-	Enumeration Area
HHID	-	Household identification variable
JICA	-	Japan International Cooperation Agency
LSMS-ISA	-	Living Standards Measurement Study-Integrated Surveys on Agriculture
MDG	-	Millennium Development Goal
MKUKUTA	-	National Strategy for Growth and Reduction of Poverty
NBS	-	Tanzania National Bureau of Statistics
THBS	-	Tanzania Household Budget Survey
TZNPS	-	Tanzania National Panel Survey
UNFPA	-	The United Nations Population Fund
UNICEF	-	The United Nations Children's Fund

Table of Contents

Contents

Background

The Tanzania National Panel Survey (TZNPS) is the first in a series of nationally representative household panel surveys that assembles information on a wide range of topics including agricultural production, non-farm income generating activities, consumption expenditures, and a wealth of other socio-economic characteristics. The first year of the survey was conducted over twelve months from October 2008 to October 2009. It was implemented by the Tanzania National Bureau of Statistics (NBS). The second wave of the TZNPS is planned for fall 2010.

The main objective of the TZNPS is to provide high-quality household-level data to the Tanzanian government and other stakeholders for monitoring poverty dynamics, tracking the progress of the Mkukuta poverty reduction strategy[1], and evaluating the impact of other major, national-level government policy initiatives. As an integrated survey covering a wide range of socioeconomic factors, it compliments other more narrowly focused survey efforts such as the Demographic and Health Survey on health, the Integrated Labour Force Survey on labour markets, the Household Budget Survey on expenditure, and the National Sample Census of Agriculture. Secondly, as a panel household survey where households are revisited over time, the TZNPS allows for the study of poverty and welfare transitions and the determinants of living standard changes, rather than only cross-sectional statistics.

NBS was advised on technical issues related to survey design and implementation by the TZNPS Technical Committee, which included representatives from line ministries, government agencies and development partners, such as the Ministry of Agriculture, Food Security and Cooperatives, Ministry of Finance, Millennium Challenge Authority Tanzania, World Bank, DFID, UNICEF, UNFPA, and JICA. The first wave of the TZNPS was supported by several donors, including the World Bank, The Bill and Melinda Gates Foundation (BMGF), UNICEF, UNFPA, and the Royal Danish Embassy, as well as the Government of Tanzania through the pooled Mkukuta funding. NBS also received management and technical support from the LSMS Team in the Development Economics Research Group (DECRG) of the World Bank.

The TZNPS is part of the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA[2]), which supports governments in seven Sub-Saharan African countries to generate nationally representative, household panel data with a strong focus on agriculture and rural development.

This document describes all aspects of the TZNPS 2008/09, including the set of survey instruments, sample design, survey implementation, and the resulting data sets.

---

[1] MKUKUTA is a Kiswahili acronym for the National Strategy for Growth and Reduction of Poverty. The current phase began in 2005 and will finish in 2010, and is the government strategy to meet the MGD and other national development goals. The focus of this round of MKUKUTA includes growth and the reduction of poverty, improved quality of life and social well-being, and governance and accountability.
[2] The Living Standards Measurement Study is an ongoing research agenda within the Development Economics Research Group of the World Bank with the goal of promoting and improving the collection of household level data in developing countries around the world. Further information can be found at www.worldbank.org/lsms.

Survey Instruments

The first wave of the TZNPS consists of several survey instruments: a Household Questionnaire, Agriculture Questionnaire, and Community Questionnaire.

The Household Questionnaire is comprised of thematic sections. A detailed description of the contents of the questionnaire can be found in Table 1. This comprehensive questionnaire allows for the construction of a full consumption-based welfare measure, permitting distributional and incidence analysis. This project also recognizes the imperative to look beyond the household as a unit of analysis in order to improve the quality, relevance and sustainability of agricultural data systems. Although data collection is structured around a household panel survey, the data on labor, education and health status were collected at the individual level. Moreover, in some household activities (like non-farm enterprise), the questionnaire records which specific members are engaged in the activity.

The questionnaires were developed in collaboration with line ministries and donor partners, including the Technical Committee, over a period of several months. The NBS solicited feedback from various stakeholders in regards to survey content and design. The questionnaires were piloted in the Morogoro region in June 2008, in conjunction with supervisor training. After piloting, the questionnaires were further revised and finalized by September 2008.Questionnaire manuals were developed with detailed instructions for field staff for use during training and as a reference over the course of the field work.

The Agricultural Questionnaire collects information relative to a household's agricultural activities. Information is collected at both the p lot and crop level on inputs, production and sales. Table 2 provides a detailed description of the contents of the questionnaire. This questionnaire was administered to any household that engaged in any farming, livestock or fishery activities.

The Community Questionnaire collects information on physical and economic infrastructure and events in surveyed communities[3], as described in Table 3. The respondents for this questionnaire vary by individual location but generally include the Village/Block Chairperson, the Village/Ward Executive officer, and several sub-village chairpeople. Information about these respondents is collected individually in section CH of community questionnaire. This questionnaire is administered by the field team supervisors in all communities included in the sample. The questionnaire section on market prices was revised mid-year to improve the formatting and, therefore, the quality of the price data. So there are two different price data sets in the community data reflecting the change in formatting of the price section.

---

[3] Note that this is not a "community" in the sociological sense, but rather a mechanism to collect information about the areas where the households selected for the survey are located. In most rural areas, EAs are defined by village boundaries and therefore community refers to the village. In urban areas the distinction is less clear, and occasionally single community questionnaires were administered to adjoining EAs.

Table 1: Household Questionnaire

Section	Name	Level of observation	Description
A-1	Household Identification	Household	household identifier variables
A-2	Survey Staff Details	Household	enumerator, supervisor, and data entry clerk identifiers, date and time of interview and data entry, and observation notes by enumerator regarding the interview
B	Household Member Roster	Individual	roster of individuals living in the household, relationship to the household, gender, year of birth, age, marital status, spouse identification, parental status, and place of birth.
C	Education	Individual	educational attainment, school characteristics, and expenditures
D	Health	Individual	general health status and on utilization of health services
E	Labor	Individual	labor market participation during the last seven days, wage work, non-farm enterprise activity, and domestic activities within the home
F	Food Outside the Household	Individual	TZ shilling value of food consumed outside the home during the last seven days
G	Children Living Elsewhere	Children living elsewhere	children of household members who live elsewhere
H	Governance	Household	government officials (if known), information on attendance at local governance meetings, religious affiliation, and participation in religious activities
I-1	Violence Against Women	Women age 15-50 years	Attitudes about violence against women and the types of violence experienced
J	Housing, Water and Sanitation	Household	dwelling characteristics, water source, sanitation facilities, and use of fuel and electricity
K	Food Consumption	Household	quantity and value of food consumed within the household during the last seven days by source.
L	Non-Food Expenditures (past one week and one month)	Household	non-food expenditure during the last week or last month
M	Non-Food Expenditures (past twelve months)	Household	non-food expenditure during the past 12 months

Section	Name	Level of observation	Description
N	Household Assets	Household	assets
O	Assistance and Groups	Household	government and non-governmental assistance received
P	Credit	Household	loans or credit received by the household during the last twelve months, conditions of the transaction, and intended use
Q	Crime and Justice	Household	crimes or disputes suffered by the household during the last twelve months, and associated interactions with the police and justice system
R	Recent Shocks	Household	economics shocks experienced by the household during the last five years
S	Deaths in the Household	Household	deaths of household members in the last two years
T-1	Household re-contact information	Household	Re-contact information
T-2	Filter Questions for Agricultural Questionnaire	Household	identification questions on households who farm or have livestock
U	Anthropometry	Individual	anthropometric information on household members aged over seven months
W	Weights	Cluster	Survey weights or expansion factors to construct nationally representative statistics

Table 2: Agriculture Questionnaire

Section	Name	Level of observation of the data file	Description
1	Household Roster	Individual	name, age, and sex of the household members, copied from the household questionnaire
2	Plot Roster	Plot	list of all plots cultivated or owned by the household during the last completed long and short rainy seasons
3A and 3B	Plot Details	Plot	detailed plot information (agricultural practices, ownership status of the land, use of fertilizers or pesticides, inputs received on credit, and labor inputs) for long and short rainy season
4A and 4B	Annual crops by Plot	Plot-crop	crops planted and harvested, seeds used, and any unexpected losses experienced for long and rainy season
5A and 5B	Annual crop Production and Sales	Crop	quantity and value of crops sold, post-harvest losses, and storage
6A and 6B	Permanent Crops by Plot	Plot-crop	age of plants, agricultural practices, quantity harvested, and unexpected losses
7	Permanent Crops – Production and Sales	Crop	quantity and value of crop sold, post-production losses, and storage
8	Outgrower Schemes and Contract Farming	Crop	outgrower or contract farming agreements for annual crops in long rainy and short rainy seasons, and permanent crops
9	Processed Agricultural Products and Agricultural Bi-Products	Crop-product	agricultural products that were processed during the last twelve months, including the cost and value at sale
10A	Livestock	Animal	livestock owned by the household during the last twelve months, caretaking practices, and unexpected losses
10B	Livestock By-Products	Animal by-product	quantity and value of livestock byproducts produced by the household during the last twelve months
11	Fishery and Aquaculture	Input expenditure	farm implements used or owned by the household during the last twelve months
12	Farm Implements and Machinery	Implement	fishing equipment owned or used by the household during the last twelve months, and inputs expended and income received from fishing practices
13	Extension	Extension source	interaction with governmental and non-governmental extension agents

Table 3: Community Questionnaire

Section	Name	Level of observation of the data file	Description
A-1	Community Identification	Community	community identifier variables and GPS coordinates
A-2	Survey Staff Details	Community	details on the enumerator team, date of interview and data-entry, and directly observed community characteristics
B	Access to Basic Services	Service	availability of basic services, the name of the nearest provider, and the distance to their location.
C	Investment Projects	Investment Project	presence of local investment projects for basic services in 2007/8, including sources and amount of funding
D	Land Use	Community	land-use practices in the village and any notable changes in land ownership/appropriation
E	Agriculture	Community	presence of and participation rate of cooperatives in the community, the availability of improved maize inputs, and an assessment of the quantity of local rainfall in the previous rainy seasons
F	Demography and Family Issues	Community	origins of the village, predominant religions and ethnic groups, and marriage and inheritance practices
G	Governance	Community	village assemblies and tribunals during the last twelve months, and their basic characteristics
H	Roster of Community Leaders	Community Leader	community leaders including their occupation, political party, and other socio-demographic characteristics
I	Crime and Policing	Community	presence of police and other security forces, and crimes that have occurred in the last month or twelve months
J	Market Prices	Item	prices for basic foodstuffs and nonfood items

Sample Design

The sample size of 3,280 households was calculated to be sufficient to produce national estimates of poverty, agricultural production and other key indicators. It will also be possible in the final analysis to produce disaggregated poverty rates for 4 different strata: Dar es Salaam, other urban areas on mainland Tanzania, rural mainland Tanzania, and Zanzibar. Alternatively, estimates of most key indicators can be produced at the zone level, as used for the Demographic and Health Survey (DHS) reports and other surveys. There are 7 of these zones in total on the mainland: North, Central, Eastern, South, Southern Highlands, West and Lake. As with any survey though, the confidence of the estimates declines as statistics are disaggregated into smaller zones. Due to the limits of the sample size it is not possible to produce reliable statistics at the regional or district level.[4]

The sample was constructed based on the National Master Sample frame which is a list of all populated enumeration areas in the country developed from the 2002 Population and Housing Census. The sample includes a partial sub-sample of households interviewed during the 2006/2007 Household Budget Survey. Sample design was done in spring of 2008.

In total, the target sample was 3,280 households in 410 Enumeration Areas (2,064 households in rural areas and 1,216 urban areas).

Implementation

The field staff was trained in Korogwe in September 2008 over four weeks, and a second training was held over three weeks after six months of field work in April 2009. The second training was necessitated by a shift in implementation strategy – from having dedicated household and agricultural enumerators to having all enumerators trained to administer both surveys. In practice, this had be the *de facto* implementation strategy for most teams since the start of the survey, and the main goal of the standardizing techniques across all teams.

Data were collected between October 2008 and October 2009. The survey was implemented by six mobile field teams, each composed of: one supervisor, three enumerators, one data entry technician, and one driver. A seventh team was added in April 2009 to relieve some of the workload from the existing teams.

The teams visited each enumeration area for between 3-4 days. The questionnaires were administered to the selected households over the course of that time. This allowed the field team to make return visits to the household to complete the entire Household questionnaire and, for farm households, Agriculture questionnaires. To ensure the depth and quality of each section of

---

[4] Description excerpted from *Tanzania National Panel Survey Report Round 1, 2008-2009* which is forthcoming from the National Bureau of Statistics. Further information on the sampling strategy will be available in this document.

the survey, the questionnaire was administered across multiple respondents to the most knowledgeable about each topic. For 25% of the sampled households, areas of all owned and/or cultivated agricultural plots were measured via GPS. Anthropometric measurements were taken for all individuals.

Data entry was done concurrently with data collection by the data entry technician, using a laptop. The data entry program was a CSPro-based system, developed by NBS with support from the World Bank. This facilitated the performance of internal crosschecks prior to departure from the enumeration area, allowing enumerators to return to households and clarify inconsistent information on the questionnaires. Data files from completed EAs were then e-mailed to headquarters using 3G modems. These files were concatenated and periodic checks were done to ensure the fieldwork was proceeding according to the calendar.

Following the completion of fieldwork and data-checks during data entry, an extensive review of data files was conducted, including verification against the paper questionnaire of outliers and missing values. Observations were returned for manual inspection of the physical questionnaires if:
- Values fell beyond five standard-deviations from the sample mean.
- Total consumption was less than the sum of purchased and home-produced consumption.
- The unit of measure was inconsistent across the questions for a single consumption item in Section K (i.e., purchased amounts were recorded in kilograms but total mounts were recorded in grams).
- Non-zero monetary expenditure on a given item was accompanied by missing values for total consumption.

When it was determined that these values were the result of data-entry error, the values were corrected. In addition, cases deemed to reflect obvious enumerator error were also corrected in this cleaning process. The majority of such cases involved the use of incorrect measurement units, e.g. recording grams as kilograms or vice versa.

Data Set

The TZNPS consists of several data files. Each data file pertains to a section of the questionnaire or a set of sections that are for the same level of observation. The complete list of data files are in the Appendix. The unique identification variables are also listed in the Appendix; there is at least one and are up to three unique identification variables in each data set.

Households are identified by a fourteen-digit number that is constructed from the district, ward, locality, enumeration area, and household number. This variable is "hhid" in the data files. When merging or linking individual data files, it is necessary to use both the household identification variable ("hhid") and the other unique identification variables pertinent to the data files being merged.

The required variables to merge data sections depend on which two sections are being merged. For example, merging sections A and J (basic identification and household characteristics)

require only the "hhid" variable as it is the single unique identifier in both datasets. Merging sections B and G (roster information with children living outside of the household) requires merging on both the hhid and the mother's id. This combination of variables will be unique in the roster datafile (as only one person can have a particular roster number), but will not be unique for the children living outside the household file (as multiple children will likely have the same mother). The merged files will be identified uniquely by the child line number ("sgchildno"), but this variable is not used in the merge itself. Similar patterns will be found in other combinations of data files.

Weighting / Expansion Factors

In order to produce nationally representative statistics with the NPS data, it is necessary to apply weighting or expansion factors. These survey weights adjust for differences in the probability of selection into the NPS sample for observations in various strata.[5]

The NPS sample is a multi-stage clustered sample design. First stage sampling involved the selection of survey clusters with the probability of selection proportional to cluster size within a stratum. In rural areas a cluster is defined as an entire village. In urban areas a cluster is defined as a census enumeration area. As a general rule, the probability of selection was higher for clusters within strata where existing data sources showed that the variance of key variables of interest for the NPS (e.g., household consumption and maize production) were likely to be very high – implying the need for more observations to produce reliable estimates. In practice, this implies that a disproportionate number of clusters were selected in urban areas including Dar es Salaam, and thus a larger weight or expansion factor is placed on rural observations in the NPS in order to reconstruct national averages.

The expansion factors can be found in the Weights data file. There are two variables of interest included: the first, "hh_weight" gives the expansion factor needed to produce nationally representative estimates. The second, "hh_weight_trimmed" is identical to the first, except that the largest and smallest 1% of observations in have been censored at the 1st or 99th percentile of ethe "hh_weight" distribution, respectively. This censoring, or "trimming" of the weights has little or no measurable effect on statistical means calculated with the NPS data, but significantly reduces the estimated standard errors or confidence intervals of sample statistics.

The Weights data file also contains unique identifiers for the first-stage sampling units ("clustered") and for the sampling strata ("strataid"). First stage sampling of clusters was stratified along two dimensions: (i) eight administrative zones (seven on Mainland Tanzania plus Zanzibar as an eighth zone), and (ii) rural versus urban clusters within each administrative zone. The combination of these two dimensions yields 16 strata.

---

[5] The details of the sample design – including the sampling strata and the use of multiple sampling frames – are discussed in a separate document, *Sample Design for the National Panel Survey*, April 2009, available from NBS upon request.

Obtaining Data

Data is available on the Tanzania National Bureau of Statistics website (www.nbs.go.tz), and may be downloaded free of charge. Inquiries sent to the LSMS team at lsms@worldbank.org will be forwarded to NBS for processing.

Appendix: TZNPS Year 1 Data Files

HOUSEHOLD QUESTIONNAIRE		
Date File	**Description**	**Unique Identifier**
SEC_A_T.dta	Household Identification (Section A), GPS (Section T)	hhid
SEC_B_C_D_E1_F_G1_U.dta	Household Roster (Section B), Education (Section C), Health (Section D), Labour (Section E), Food Consumption Outside Household (Section F), Anthropometrics (Section U)	hhid, sbmemno
SEC_E2.dta	Additional labour enterprises (Sub section of E, q24-44)	hhid, namba_ya_biashara
SEC_G2.dta	Children Living Elsewhere (Section G)	hhid, sgchildno
SEC_H1_J_K2_O2_P1_Q1_S1.dta	Housing, Water and Sanitation (Section J), Governance (Section H, q1-3, 10,11), Food Consumption (Section K, q7-9), Assistance and Groups (Section O, q6), Partial Crime and Justice (Section Q, q1- 3, q16-20)	hhid
SEC_H2.dta	Governance (Section H, q4-7)	hhid
SEC_H3.dta	Governance (Section H, q8, 9)	hhid, shmeet
SEC_I.dta	Violence Against Women (Section I)	hhid, siq1
SEC_K1.dta	Food Consumption (Section K)	hhid, skcode

SEC_L.dta	Non-Food Expenditures (Section L)	hhid, slcode
SEC_M.dta	Non-Food Expenditures (Section M)	hhid, smcode
SEC_N.dta	Household Assets (Section N)	hhid, sncode
SEC_O1.dta	Assistance and Groups (Section O, q1-5)	hhid, socode
SEC_O3.dta	Assistance and Groups (Section O, q7-15)	hhid, somemno
SEC_P2.dta	Credit (Section P)	hhid, sp2code
SEC_Q2.dta	Crime and Justice (Section Q, q4-25)	hhid, sq2code
SEC_Q3.dta	Crime and Justice (Section Q, q21-28)	hhid, sqq21
SEC_R.dta	Recent Shocks (Section R)	hhid, srcode
SEC_S2.dta	Deaths in the Household (Section S)	hhid, sscode

AGRICULTURE QUESTIONNAIRE		
Date File	**Description**	**Unique identifier**
SEC_1_ALL.dta	Household Roster (Section 1)	hhid, rosterid
SEC_2A.dta	Plot Roster (Section 2A)	hhid, plotnum
SEC_2B.dta	Plot Roster (Section 2B)	hhid, plotnum

SEC_3A.dta	Plot Details (Section 3A)	hhid, plotnum
SEC_3B.dta	Plot Details (Section 3B)	hhid, plotnum
SEC_4A.dta	Crops by Plot (Section 4A)	hhid, plotnum, zaocode
SEC_4B.dta	Crops by Plot (Section 4B)	hhid, plotnum, zaocode
SEC_5A.dta	Crops – Household Totals (Section 5A)	hhid, zaocode
SEC_5B.dta	Crops – Household Totals (Section 5B)	hhid, zaocode
SEC_6A.dta	Permanent Crops by Plot (Section 6A)	hhid, plotnum, zaocode
SEC_6B.dta	Permanent Crops by Plot (Section 6B)	hhid, plotnum, zaocode
SEC_7A.dta	Permanent Crops by Crop (Section 7A)	hhid, zaocode
SEC_7B.dta	Permanent Crops by Crop (Section 7B)	hhid, zaocode
SEC_8A.dta	Outgrower Scheme and Contract Farming (Section 8A)	hhid, zaocode
SEC_8B.dta	Outgrower Scheme and Contract Farming (Section 8B)	hhid, zaocode
SEC_8C.dta	Outgrower Scheme and Contract Farming (Section 8C)	hhid, zaocode
SEC_9_ALL.dta	Processed Agricultural Products and Agricultural Bi-Products (Section 9)	hhid, zaocode, byproduct
SEC_10A.dta	Livestock (Section 10A)	hhid, animal

SEC_10B.dta	Livestock By-Products (Section 10B)	hhid, lvstkprod
SEC_11_ALL.dta	Farm Implements and Machinery (Section 11)	hhid, impcode
SEC_12A.dta	Partial Fishery and Aquaculture (Section 12, q2-9)	hhid, itemcode
SEC_12B.dta	Partial Fishery and Aquaculture (Section 12, q10-15)	hhid, itemcode
SEC_12C.dta	Partial Fishery and Aquaculture (Section 12, q24-26)	hhid, areacode
SEC_12D.dta	Fishery and Aquaculture (Section 12, q16-20)	hhid, itemcode
SEC_13A.dta	Extension (Section 13, q1-6)	hhid, source
SEC_13B.dta	Extension (Section 13, q7-9)	hhid, source
SEC_QNFLOW.dta	Household Identification (Section 1A), Plot Roster (Section 2, q1, 5, 6), Plot Details (Section 3, q1), Permanent Crop by Plot (Section 7, q1), Outgrower (Section 8, q1), Processed Agricultural Products (Section 9, q1), Livestock (Section 10A, q1), Farm Implements (Section 11, q1), Fishery and Aquaculture (Section 12, q21-23), Extension (Section 13, q11-15)	hhid
SEC_PLOTGPS.dta	GPS Data	hhid
SEC_NETWORK.dta	Network Roster Card	hhid, nid

COMMUNITY QUESTIONNAIRE		
Date File	**Description**	**Unique Identifier**
SECTA1A2.dta	Community Identification (Section A)	Community
SECTCB.dta	Access to Basic Services (Section B)	Service
SECTCC.dta	Investment Projects (Section C)	Investment Project
SECTCD.dta	Land Use (Section D)	Community
SECTCEFG.dta	Agriculture (Section E), Demography and Family Issues (Section F), Governance (Section G)	Community
SECTCH.dta	Roster of Community Leaders (Section H)	Community Leader
SECTCI.dta	Crime and Policing (Section I)	Community
SECTCJ.dta	Market Prices (Section J)	Item