NATIONAL BUREAU OF STATISTICS

Basic Information Document

National Panel Survey

UNIFORM PANEL DATASET

2019

[UNITED REPUBLIC OF TANZANIA]

ACRONYMS

BID	-	Basic Information Document
FYDPII	-	Five Year Development Plan II
HBS	-	Household Budget Survey
HHID	-	Household Identification
ILFS	-	Integrated Labour Force Survey
LSMS	-	Living Standards Measurement Survey
MDG	-	Millennium Development Goal
NBS	-	National Bureau of Statistics
NPS	-	National Panel Survey
NSCA	-	National Census Sample of Agriculture
PHC	-	Population and Housing Census
UPHI	-	Uniform Panel Household Identifier
UPI	-	Uniform Panel Identifier
UPD	-	Uniform Panel Dataset

Table of Contents

Background 4
Survey Instrument 5
Variable Names: 6
Response Codes: 6
Changes in Questions Over Rounds: 6
Changes in Questions Over Rounds: 7
Data Sets 7
Panel Sequence: 8
Household Panel Sequences: 8
Extended Panel Sample: 9
Refresh Panel Sample: 10
Initial Sample (R1-R3): 10
Split-off Households: 10
Data Set Layout: 12
Naming Convention for "r_hhid": 12
Merging Between Data Sets: 13
Obtaining the Data: 14
UPD Module Information 15
Appendix 29
Appendix I: NPS UPD Naming Convention 29
Appendix II: Questionnaire Footnotes 30

Background

The Tanzania National Panel Survey (NPS) is a series of nationally representative multi-topic household surveys that collect information on topics including agricultural production, non-farm income generating activities, consumption expenditures, and a wealth of other socio-economic characteristics. The first NPS was implemented in 2008/09, with additional rounds implemented in 2010/11, 2012/13, 2014/15, and 2019/2020 (currently underway).

The main objective of the NPS is to provide high-quality household-level data to government and other stakeholders for monitoring poverty dynamics, tracking the progress of national poverty reduction strategies such as the Five Year Development Plan (FYDP) II and MKUKUTA II[1], and to evaluate the impact of other major, national-level government policy initiatives. As an integrated survey covering a number of different socioeconomic factors, it compliments other more narrowly focused survey efforts, such as the Demographic and Health Survey on health, the Integrated Labour Force Survey (ILFS) on labour markets, the Household Budget Survey (HBS) on expenditure, and the National Sample Census of Agriculture (NSCA).

Panel data possess several advantages over conventional cross-sectional and time-series data, including their power to better isolate the effects of specific actions, treatments, or policies. While panel surveys offer the potential for more robust analysis of the topic at hand, they may also be difficult to work with due to time- and situation-driven variations in the tools and techniques used in administration of the surveys. Though survey variations over time are necessary and generally unavoidable, inconsistencies between panel datasets may increase both the difficulty and cost of use. When such variations are mitigated, the ease of use increases and subsequently the overall value of the panel survey increases.

The NPS Uniform Panel Dataset (UPD) is a collection of topic-specific datasets, meticulously aligned and engineered with the aim of facilitating the use of and improving access to the wealth of panel data offered by the NPS. The design of the NPS-UPD combines the four completed rounds of the NPS – NPS 2008/09 (R1), NPS 2010/11 (R2), NPS 2012/13 (R3), and NPS 2014/15 (R4)[2] – into pooled, module-specific datasets. The supporting panel survey instruments offer the ease of comparability over time, with modifications and variances easily identifiable as well as those aspects of the questionnaire which have remained identical and offer consistent information. By providing all module-specific data over time within compact, pooled datasets, the UPD eliminates the need for user-generated merges between rounds and present data in a clear, logical format, increasing both the usability and comprehension of complex panel data.

[1] MKUKUTA II is a Kiswahili acronym for the National Strategy for Growth and Reduction of Poverty. Operating between 2010 and 2015, MKUKUTA II is the government strategy to meet the MDGs and other national development goals. The focus of MKUKUTA II includes growth and the reduction of poverty, improved quality of life and social well-being, and governance and accountability.

[2] For simplicity, NPS rounds will be hereafter be referred to as the following: “R1” (NPS 2008/09), “R2” (NPS 2010/11), “R3” (NPS 2012/13), and “R4” (NPS 2014/15).

This Basic Information Document (BID) provides a brief overview of the Tanzania National Panel Survey (NPS), but focuses primarily on the theoretical development and application of panel data, as well as key elements of the Uniform Panel Dataset (UPD) and survey instrument generated by the four completed rounds of the NPS. As this BID does not describe in detail the background, development, or use of the NPS itself, the round-specific NPS BIDs should be used to supplement the information provided here.

Survey Instrument

The NPS-UPD survey instrument highlights the comparability of both questions and response codes across rounds and allows for easy identification of information reported in all rounds of the NPS. There are 26 module-specific questionnaires which correlate directly to the NPS-UPD datasets in terms of module names, variable names, variable labels, and response codes.

The format of the NPS-UPD survey instrument is similar to previously disseminated NPS survey instruments. Each module has a questionnaire and clearly identifies if the module collects information at the individual or household level. Within each module-specific questionnaire of the NPS-UPD survey instrument, there are five distinct sections, arranged vertically: (1) the UPD - "U" on the survey instrument, (2) R4, (3), R3, (4) R2, and (5) R1 – the latter 4 sections presenting each questionnaire in its original form at time of its respective dissemination.

The uppermost section of each module's questionnaire ("U") represents the model *uniform panel questionnaire*, with questions generated from the comprehensive listing of questions across all four completed rounds of the NPS and codes generated from the comprehensive collection of codes. The following sections are arranged vertically by round, considering R4 as most recent. While not all rounds will have data reported for each question in the UPD and not each question will have reports for each of the UPD response codes listed, the NPS-UPD survey instrument represents the visual, all-inclusive set of information collected by the NPS over time.

The four round-specific sections (R4, R3, R2, R1) are aligned with their UPD-equivalent question, visually presenting their contribution to compatibility with the UPD. Each round-specific section includes the original round-specific variable names, response codes and skip patterns (corresponding to their respective round-specific NPS datasets, and despite their variance from other rounds or from the comprehensive UPD code listing)[3].

[3] Skip patterns within the round-specific sections of questionnaires are directed to the variable name of the UPD to facilitate ease of direction, as more complicated modules often cannot maintain the conventional numerical order of variables within each of the round (R4 is again considered most recent and therefore is best preserved as the conventional structure).

Variable Names:

Variable names for UPD questions were assigned following the comprehensive listing of questions from all four rounds within each module and used the naming convention defined in Table 2. Therefore, the UPD variable name will not always match the original round-specific variable name.

Table 2. UPD Variable Naming Convention

EXAMPLE 1: No Subset	**EXAMPLE 2: Subset Question**
Variable Name: **hb_01**	Variable Name: **hi_34_1**
where *h* = household *b* = module B *01* = question number one	*where* *h* = household *i* = module i *34* = question number 34 *_1* = subset question of question 34

Response Codes:

Response codes for each UPD question were generated from the comprehensive listing of response codes from all four NPS rounds for that question. Similar to the variable naming structure, since response code values for the UPD questions were re-assigned following this comprehensive listing they will not necessarily correspond to each original round-specific response code value. The UPD code listings intend to present the all-inclusive set of information gathered at any point in time by the NPS.

Changes in Questions Over Rounds:

Identical Questions: Whenever possible, questions that are identical or nearly identical across rounds (i.e. "What kind of toilet facility do members of your household generally use?" and "What is the main toilet facilities usually used in this household?") are presented as a single, unified UPD question. These will be vertically aligned in the module's questionnaire. Minor changes in question wording that are not anticipated to affect interpretation or results are noted as footnotes, but the slight variations will not be presented as unique UPD questions.

Distinct Variations of Questions: When a question is considerably different across rounds (i.e. "What measures does this household take to ensure the safety of drinking water?" and "What do you usually do to the water to make it safer to drink in the dry season?"), all variations of the question will be included in the UPD questionnaire with the appropriate blackouts for the respective rounds that the UPD question is not present in. For these cases, it will be up to the user's discretion to combine.

Prevalence of Questions: Whether a question exists in some or all of the four NPS rounds, it will be included in the UPD questionnaire and will either have the respective round-specific variables presented below it, or will have appropriate blackouts for the respective rounds that the question is not present in.

Changes in Questions Over Rounds:

Identical Response Codes: Whenever possible, response codes that are identical or nearly identical (i.e. "OTHER" and "OTHER, SPECIFY" or "BOIL" and "BOILING WATER") across rounds are presented as a single, unified UPD response code. Minor changes in wording that are not anticipated to affect interpretation or results are noted as footnotes, but the slight variations will not be presented as unique UPD response codes.

Distinct Variations of Response Codes: When a response code is considerably different across rounds (i.e. "RELIGIOUS" and "RELIGIOUS – BOARDING" or "PIT LATRINE", "PIT LATRINE WITH SLAB (WASHABLE)" and "PIT LATRINE WITH SLAB (NOT WASHABLE)"), all variations of the response code will be included in the UPD response code listing. For these cases, it will be up to the user's discretion to combine.

Prevalence of Response Codes: Whether a response code exists in some or all of the four NPS rounds, it will be included in the UPD code listing and will be identified in the round-specific sections by a box. If additional information is relevant, a footnote will accompany the code. In these instances, the values for codes may differ from the original, round-specific codes to accommodate for more, less, or different codes. These differences will also be identified in the round-specific sections by a box.

Data Sets

The NPS-UPD was developed in accordance with the NPS-UPD survey instrument, and aims to neatly present and facilitate the use of panel data collected by the NPS. By offering compact, standardized data sets, data are readily available and pre-standardized for optimal panel analyses.

The design of the uniform panel dataset combines the four rounds of NPS data – NPS 2008/09 (R1), NPS 2010/11 (R2), NPS 2012/13 (R3), and NPS 2014/15 (R4) – into pooled, module-specific datasets. The NPS-UPD consists of 41 data files, each pertaining to a single module or portion of the module (i.e. "UPD4_HH_Q1" and "UPD4_HH_Q2"). Each data file includes data for all rounds that reported information on that module. As mentioned, modules will occasionally be split into multiple datasets if the level of observation is not consistent within the module. The complete set of data files, as well as the level of observation, unique identifier for that module, and rounds of data included, are described in the UPD Module Information tables.

SAMPLE DESIGN

While the same sample of respondents was maintained over the first three rounds of the NPS, longitudinal surveys tend to suffer from bias introduced by households leaving the survey over time; i.e. attrition. Although the NPS maintains a highly successful recapture rate (roughly 96% retention at the household level), minimizing the escalation of this selection bias, a refresh of longitudinal cohorts was done for the NPS 2014/15 to ensure proper representativeness of estimates while maintaining a sufficient primary sample to maintain cohesion within panel analysis. A newly completed Population and Housing Census (PHC) in 2012, providing updated population figures along with changes in administrative boundaries, emboldened the opportunity to realign the NPS sample and abate collective bias potentially introduced through attrition.

To maintain the panel model of the NPS, the sample design for NPS 2014/15 consisted of a combination of the original NPS sample and a new NPS sample. A nationally representative sub-sample was selected to continue as part of the "Extended Panel" while an entirely new sample, "Refresh Panel", was selected to represent national and sub-national domains. Similar to the original sample in NPS 2008/09, the sample design for the "Refresh Panel" also allows for analysis at four primary domains of inference, namely: Dar es Salaam, other urban areas on mainland Tanzania, rural mainland Tanzania, and Zanzibar. The new cohort in NPS 2014/15 will be maintained and tracked in all future rounds between national censuses or until refreshed again.

Panel Sequence:

Not every NPS household or individual will have data present for all four rounds. In each round, circumstances such as tracking success, attrition through loss or refusal, and the revised sampling design in R4 contribute to a variety of *panel sequences* that represent distinct combinations of movement and participation over time. As both households and individuals are tracked over time, both *household panel sequences* and *individual panel sequences* are formed. These sequences correspond to the data sets of the same information level i.e. household panel sequences are observed in household-level datasets (e.g. education) and individual panel sequences are observed in individual-level datasets (e.g. housing, water, and sanitation).

Household Panel Sequences:

Each household panel sequence represents a unique combination of household movement and participation over time. Sequences are composed of "bands" of data, which are determined by a successful interview and thus the presence of data in a particular round.

Figure 1A.

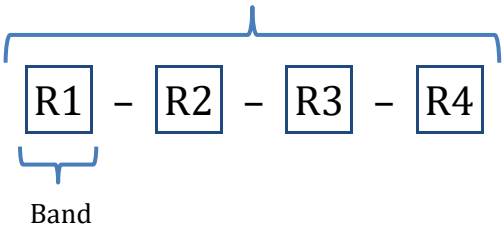
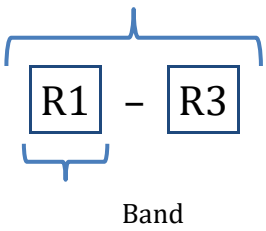


Figure 1B.



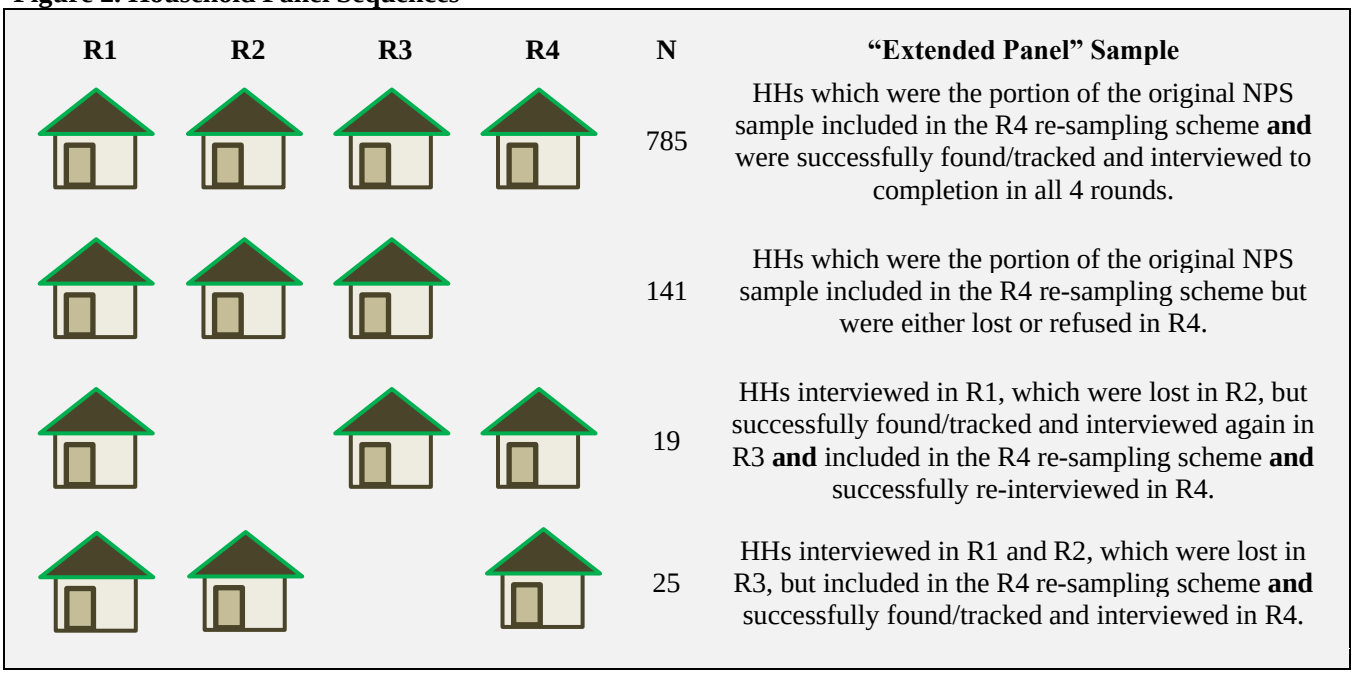
The concept of these sequences is integral for understanding the architecture of the universal panel datasets. While the following illustrations visually present "gaps" in data (i.e. from a lost household or refusal), no gap or placeholder is included in either the sequences or in the universal panel datasets (Figure 1B); sequences are compiled only of existing bands of data. While sequences differ within the "Extended Panel Sample", "Refresh Panel Sample", and "Initial Sample (R1 – R3)", the eight sequences below create the universe of household panel sequences observable in the household-level universal panel datasets.

Extended Panel Sample:

From the original NPS sample, a nationally representative sub-sample was selected to continue as the "Extended Panel" sample, with data collection attempts made in each of the four rounds. Of these, 785 households were successfully found/tracked and interviewed in each of R1, R2, R3, and R4 (Figure 2, Example 1), regardless of individual member movement in or out of the household (as long as an original member was continuously present in the household for each successive round) or the physical movement from one structure to another structure.

The *household panel sequence* for the extended panel sample is thus generated from the progression of the originating household in R1 through R4 and consists of a maximum of four possible bands of data. In very few instances (n = 19), a household was lost in R2 but found/tracked in the subsequent rounds and re-integrated into the panel (Figure 2, Example 3). As the revised sampling design in R4 was derived from the original sample, there will also be a panel sequence of R1-R2-R4 in the universal datasets (Figure 2, Example 4), representing those households interviewed in R1 and R2, lost in R3, and successfully found/tracked in R4 and interviewed (n = 25).

Figure 2. Household Panel Sequences



Refresh Panel Sample:

As households in the "Refresh Panel" sample were interviewed for the first time in R4, the *household panel sequence* for these households is simply the single band of data available in R4 (Figure 3).

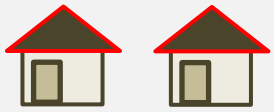
Figure 3. Household Panel Sequences

R1	R2	R3	R4	N	"Refresh Panel" Sample
				3,352	HHs included in the R4 re-sampling scheme for the first time in R4 **and** successfully interviewed; these households will have no information prior to R4.

Initial Sample (R1-R3):

The remaining sequences are for households interviewed under the initial sampling scheme (i.e. households *excluded* in the R4 revised sampling scheme). The *household panel sequence* for these households is thus generated from the progression of the original household in R1 through the following two rounds (Figure 4).

Figure 4. Initial Sample (R1-R3) Sequences

R1	R2	R3	R4	Initial Sample (R1 – R3)
			N/A	HHs excluded in the R4 re-sampling scheme which were successfully found/tracked and interviewed in all 3 eligible rounds (R1, R2, R3).
			N/A	HHs excluded in the R4 re-sampling scheme which were successfully found/tracked and interviewed in R1 & R2 and either lost or refused in R3 and unable to be interviewed.
			N/A	HHs excluded in the R4 re-sampling scheme which were successfully interviewed in R1 and either refused in R2 or were lost in the following round(s) and unable to be interviewed.
			N/A	HHs excluded in the R4 re-sampling scheme which were successfully interviewed in R1, lost in R2, but successfully found/tracked again in R3 and re-interviewed.

Split-off Households:

When a member(s) moves away from their original NPS household they form a "split-off" household, which is also tracked and interviewed as a part of the NPS sample. Household panel sequences must include an origination household, therefore the household panel sequence for split-off households is generated from a combination of the original household(s) and the subsequent split-off household(s). In other words, while R2-R3-R4 is technically a sequence of movement for a new *member* in a split-off household (i.e. a spouse that marries an original NPS member and now lives with them in a different

household), it does not characterize a *household* panel sequence – the R1 household of the original NPS member now in the split-off household with the new member(s) would be included to create the household panel sequence (Figure 5, Example 1).

As a split-off can occur at any point in time, these households exist in both the extended panel sample and the initial sample. Since household panel sequences serve as a representation of movement and participation and not household make-up, it is important to note that the household panel sequences for a split-off household **do not** differ from the sequences for an original NPS household. Figure 5 and 6 illustrate two examples of a split-off household and the identical sequence for an original NPS household.

Figure 5. Split-Off and Original Household in "Extended Panel" Sample

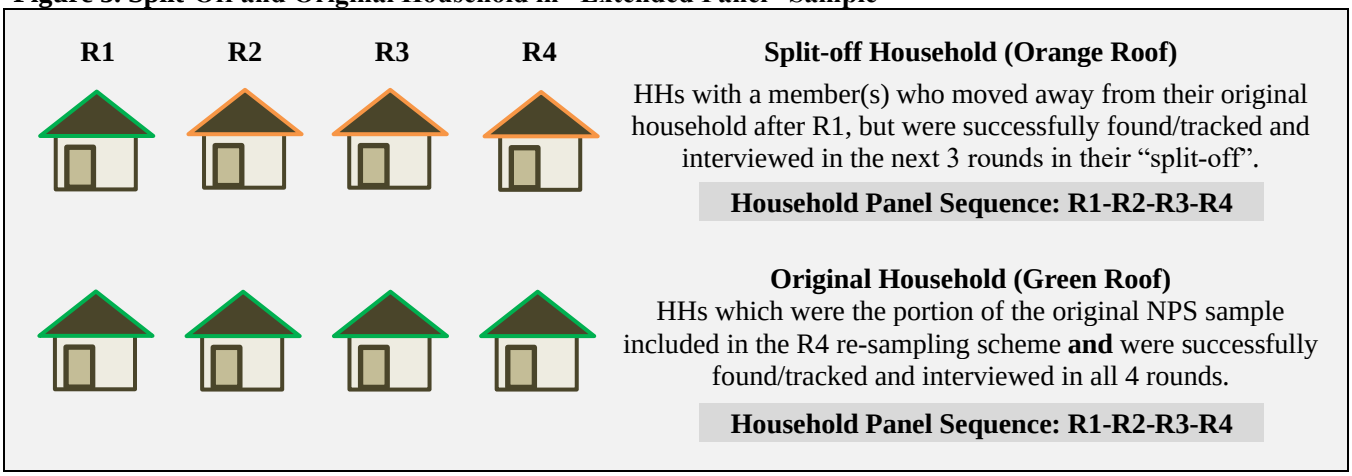
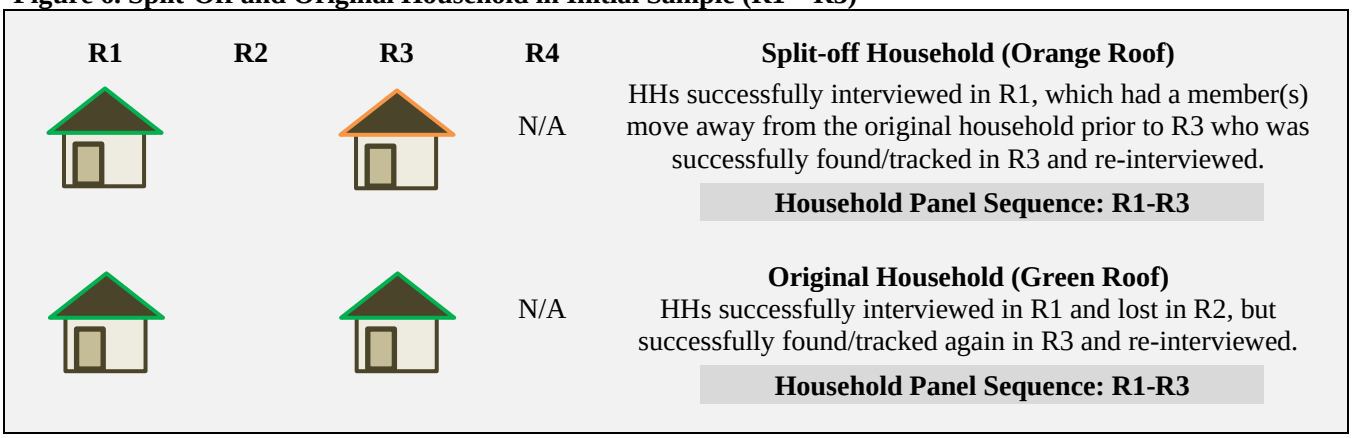


Figure 6. Split-Off and Original Household in Initial Sample (R1 – R3)



Data Set Layout:

The NPS-UPD datasets are arranged in a *long* format, sorted by a universal identifier (either at the individual or household level) and then by round (Table 3). Two variables were generated during development of the UPD datasets to uniquely identify panel sequences after combining all four rounds of data: the Uniform Panel Identifier (UPI) for individuals and the Uniform Panel Household Identifier (UPHI).

The following variables shape the foundation of the NPS-UPD datasets:

- *UPI:* unique identifier for each individual panel sequence
- *UPHI*: unique identifier for each household panel sequence
- *round*: numeric variable with values = 1 - 4 corresponding to the four rounds of NPS data; for panel sequences missing information (i.e. lost or refused) there will be no placeholder, the data line will simply not exist (Example: Table 3, *UPHI = 336* only presents data for R1 and R2.
- *r_hhid*: full 14-, 16-, or 7-digit HHID, corresponding to the round per data line

Module-level variables (i.e. "hi_01") in the UPD data sets utilize a combination of "h" for household and the lowercase module letter (i.e. "i") as the prefix to distinguish them from round-specific data releases. Again, the question number in the UPD variable name will not necessarily match the number of their round-specific variable; however, they will correspond directly to the UPD variable names in the NPS-UPD survey instrument.

Table 3. Stata Dataset Excerpt – Module_I – Housing, Water, & Sanitation

	UPHI	round	r_hhid	hi_01	hi_02
899	334	1	53020150090067	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
900	334	2	5302015009006701	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
901	334	3	3961-001	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
902	335	1	53020150090067	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
903	335	2	5302015009006703	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
904	335	3	3692-001	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
905	336	1	53020150090065	OWNER OCCUPIED	TITLE DEED FOR LAND
906	336	2	5302015009006501	OWNER OCCUPIED	TITLE DEED FOR LAND

Naming Convention for "r_hhid":

Please see round-specific NPS BIDs for detailed explanations on the naming conventions of household IDs (HHID). The round-specific HHID variables are known as the following in their individual releases: "hhid" (R1), "y2_hhid" (R2), "y3_hhid" (R3), and "y4_hhid" (R4). These have been condensed under the single variable "r_hhid" in the UPD data sets and are distinguishable only in combination with "*round"*.

For both originating and split-off households, the HHID in post-R1 rounds is generated using the lowest two-digit roster ID (e.g. 1=01; 3=03; 12=12) of the original member(s) in the household in the subsequent round. The other digits vary between rounds (i.e. R1 and R2 were longer numeric strings which were then minimized in R3 and maintained, or recreated in R4 for extended and refresh panel households, respectively).

As data layout and round labeling is identical between originating and split-off households, the latter is identified only by their HHID, whose last two digits in the split-off round will be the lowest roster ID of the split-off member(s) and a value other than "01" (Example: Table 3, UPHI = 335, round = 2).

Figure 6 illustrates a member(s) who has moved from their original household and returned to the NPS sample in a non-sequential round (i.e. R1 and again in R3). While all previously stated panel concepts remain, the last three digits of the split-off HHID in R3 are in the 100's to indicate **uncollected data** in R2 (Table 4).

Table 4. Stata Dataset Excerpt – HH_Module_I – Housing, Water, & Sanitation

	UPHI	round	r_hhid	hi_01	hi_02
11	5	1	55020180210068	OWNER OCCUPIED	NO DOCUMENTATION AT ALL
12	5	3	3922-108	OWNER OCCUPIED	LAND SALE AGREEMENT

Merging Between Data Sets:

The design of the UPD eliminates the need for user-generated merges *across* rounds for a particular module; however, merges between different modules will still occur (i.e. merging an individual's roster and education information, or food consumption and household sanitation information. The individual identification variable is the Uniform Panel Identifier ("UPI") and the household identification variable is Uniform Panel Household Identifier ("UPHI") in the data files.

The panel nature of these datasets means that most merging is simplified to a single, unique variable describing either an "individual" or a "household", either **UPI** or **UPHI**, in accompaniment with **round**. *Note*: *See previous section for a description of the generation of these identifying variables.* When merging or linking most individual-level data files, it is necessary to use only the individual identification variable, **UPI**, and round (as each individual with that UPI will have between 1 - 4 data lines, each representing a round). When merging or linking most household-level data files, it is necessary to use only the household identification variable, **UPHI**, and **round**[4]. All household level sections have been harmonized to provide a consistent number of observations in each module across like units of observation.

Additional variables required to merge data across sections depend on the particular sections being merged. In modules with multiple line items items (i.e. Module J1: Consumption of Food Over the Past One Week, or Module P: Credit, etc.), the specific item code will need to be included with the combination of **UPHI** and **round**. The unique identifying combination of variables in each module can be found in the module-specific tables of the UPD Module Information section.

[4] The **r_hhid** variable can be substituted for the **round** variable as the unique identifier. See UPD Module Information (page 15) for exact unique identifiers in each module and all possible combinations.

REVISIONS TO DISSEMINATED NPS DATA

Two households in NPS 2012/13 were assigned an incorrect "y3_hhid" during dissemination of NPS 2012/13 data. These have been corrected in the UPD datasets but may result in imperfect merges if using past data.

CASE 1: y3_hhid = **"1875-105"** was corrected to y3_hhid = **"1877-001"**
CASE 2: y3_hhid = **"2431-105"** was corrected to y3_hhid = **"2432-006"**

Obtaining the Data:

Data sets for the NPS-UPD are available free of charge on the Tanzania National Bureau of Statistics website (www.nbs.go.tz), as well as the World Bank LSMS-ISA website (www.worldbank/lsms-isa), and the World Bank Microdata Library (microdata.worldbank.org). Inquiries pertaining to the data may be sent to the LSMS team at lsms@worldbank.org.

UPD Module Information

MODULE B: HOUSEHOLD MEMBER ROSTER			
Level of Observation: **Unique Identifier(s):**	Individual UPI, round **or** UPI, r_hhid	**UPD Data File:** **Rounds Present:**	UPD4_HH_B.dta R1, R2, R3, R4

Description:
Roster of household members, individual characteristics including: sex, age, relationship to the household head, panel member identification, presence in household, general occupation, parental status, place of birth, marital status, and spouse identification.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta **R3:** HH_SEC_B.dta
R2: HH_SEC_B.dta **R4:** HH_SEC_B.dta

Key Notes:
For UPD question **hbq06**: if an individual was not present in the previous round, the enumerator is instructed to record "99". In "**UPD_HH_B.dta**," value labels for this variable were adjusted to distinctly represent what round they were not present in, where "99" = "NOT SURVEYED IN HBS", "999" = "NOT PRESENT IN Y1", "9999" = "NOT PRESENT IN Y2", and "99999" = NOT PRESENT IN Y3".

MODULE C: EDUCATION			
Level of Observation: **Unique Identifier(s):**	Individual UPI, round **or** UPI, r_hhid	**UPD Data File:** **Rounds Present:**	UPD4_HH_C.dta R1, R2, R3, R4

Description:
Educational enrollment and attainment; literacy, education history/level attainment, current enrollment, school characteristics, temporary absence, education examination (participation and score) and education expenditures.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta **R3:** HH_SEC_C.dta
R2: HH_SEC_C.dta **R4:** HH_SEC_C.dta

Key Notes:
None

MODULE D: HEALTH

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_D.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
General health status and utilization of health services; source and financing of health treatments / hospitalization, disaggregated health expenditures, disability, bed net use, pregnancy, prenatal care and births, child health and ailments / diarrhea.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta **R3:** HH_SEC_D.dta
R2: HH_SEC_D.dta **R4:** HH_SEC_D.dta

Key Notes:
In general, the NPS advises the household head or the respondent most knowledgeable about a particular module to respond to questions asked. Even in the case of individual-level modules (i.e. health), this method facilitates the most consistent and accurate reporting. In very few cases, this instruction differed across rounds, altering the terms used in the questions themselves. In R1, respondents 12 years and older were instructed to answer for themselves, so R1 questions are worded with respect to the individual answering the question, i.e. "Have you visited a health care provider…". In R2, R3, and R4, the household head/most knowledgeable respondent was instructed to answer on behalf of the individual, and thus the questions are worded as follows: “Has [NAME] visited a health care provider…”. The UPD questionnaire uses the latter terminology as these slight variations in wording are not anticipated to affect the interpretation. A footnote is provided in Module D of the questionnaire for the first instance but this will apply to the remainder of questions in the R1 questionnaire.

MODULE E: LABOUR

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_E.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
Labour market participation; activities including unpaid apprenticeship, wage work, non-farm enterprise, and agricultural activity (including livestock and fishery activities) in the last seven days and in the last twelve months. Also includes unemployment and steps taken to find work, secondary activities, industry (TASCO) and occupation (ISIC) codes, wages earned, contract type, employer provide safety nets, and general domestic activities.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta **R3:** HH_SEC_E.dta
R2: HH_SEC_E1.dta **R4:** HH_SEC_E.dta

Key Notes:
In general, the NPS advises the household head or occasionally the respondent most knowledgeable about a particular module to respond to the questions asked by the enumerator. Even in the case of individual-level modules (i.e. education, health), this facilitates the most consistent and accurate reporting. In very few cases, this instruction differed across rounds, thus altering the terms used in the questions themselves.

In R2, questions were worded in respect to the individual answering the question on behalf of themselves, i.e. "Did you do any work of any type…". In R1, R3, and R4, the pronoun "[NAME]" was used in all questions and thus is the format used in the text of the UPD questionnaire. Again, a footnote is provided in Module E of the questionnaire for the first instance but this will apply to the remainder of questions in the R2 questionnaire.

In R1, questions in the self-employment section used the pronoun "you"/"your" instead of "[NAME]", similar to the language used in R2. The text of UPD questions with this footnote is identical to the text in the questions in both R1 and R2 and is meant only to highlight the deviation from the language used throughout the remainder of R1.

MODULE F: FOOD CONSUMED OUTSIDE THE HOUSEHOLD

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_F.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
Value of food consumed outside the home during the last seven days.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta
R2: HH_SEC_F.dta
R3: HH_SEC_F.dta
R4: HH_SEC_F.dta

Key Notes:
None

MODULE G: SUBJECTIVE WELFARE & CRIME

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_G.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid	**Rounds Present:**	R1 *(partial),* R2, R3, R4

Description:
Self-reported level of satisfaction with health, financial status, housing, job, services, and safety. Also includes perceived status at present, and as of 2 years prior to the survey round. This section also collects information about any crime that the household may have been a victim of in the 12 months prior to the survey.

Original Data Files:

R1: Not present
R2: HH_SEC_G.dta
R3: HH_SEC_G.dta
R4: HH_SEC_G.dta

Key Notes:
No specific module was dedicated to Subjective Welfare & Crime in R1; however, Module E of R1 included the question regarding level of satisfaction with various components of life, which is found in the Subjective Welfare & Crime module of R2, R3, and R4. In the UPD survey instrument and UPD datasets, this question has been removed from Module E (SEC_B_C_D_E1_F_G1_U.dta) and provided in Module G. This question can be identified in the UPD Module G survey instrument by the R1 variable name (i.e. **seq49_1**, aligned with UPD variable **he_134_1**).

SECTION H: FOOD SECURITY

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_H.dta
Unique Identifier(s):	UPHI, round **or** UPHI, r_hhid	**Rounds Present:**	R2, R3, R4

Description:
Information on the household's diet and food intake patterns, months of food insecurity, and distribution of food within the household.

Original Data Files:

R1: Not present
R2: HH_SEC_I1.dta, HH_SEC_I2.dta
R3: HH_SEC_H.dta
R4: HH_SEC_H.dta

Key Notes:
For explanation of data derivation only: the original R2 NPS datasets separated the monthly incidence question (**hh_i09_1_01 – hh_i09_3_12**) from the rest of the Food Security module. However, since all Food Security data were configured in a wide format, the two R2 datasets easily translated into the UPD datasets.

SECTION I1: HOUSING, WATER, & SANITATION

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_I1.dta
Unique Identifier(s):	UPHI, round **or** UPHI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
Tenure status, rental expenditure, estimated rental value, dwelling maintenance and improvement expenditures, dwelling characteristics, including size, construction materials, toilet facilities, and means of garbage disposal. Also includes main type of cooking and lighting fuel, access to water/drinking water, and time spent collecting water. The questions on water/drinking water are asked separately for the rainy and dry season.

Original Data Files:

R1: SEC_H1_J_K2_O2_P1_Q1_S1.dta **R3:** HH_SEC_I.dta
R2: HH_SEC_J1.dta **R4:** HH_SEC_I.dta

Key Notes:
None

SECTION I2: HOUSING, WATER, & SANITATION

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_I2.dta
Unique Identifier(s):	UPHI, round, itemid **or** UPHI, r_hhid, itemid	**Rounds Present:**	R2

Description:
Additional detailed information on water sources reported, including type of container used for collection, amount and frequency of collection, and average price if purchased.

Original Data Files:

R1: Not present **R3:** Not present
R2: HH_SEC_J2.dta **R4:** Not present

Key Notes:
None

SECTION J1: CONSUMPTION OF FOOD OVER THE PAST ONE WEEK

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_J1_1-8.dta
Unique Identifier(s):	UPHI, round, hj_00 **or** UPHI, r_hhid, hj_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Consumption of 61 key items/ item groups over the last 7 days. Quantity and value of consumed items from purchases, own-production and gifts are recorded.

Original Data Files:

R1:	SEC_K1.dta	**R3:**	HH_SEC_J1.dta
R2:	HH_SEC_K1.dta	**R4:**	HH_SEC_J1.dta

Key Notes:
In R1-R3, item **hj_00** = 108 was the only item code for cereal products (item **hj_00** = 108 represented "Wheat, barley, grain, and other cereals"). In R4, this item was parsed into two distinct item codes: **hj_00** = 1081 ("Wheat flour"), and **hj_00** = 1082 ("Barley, grain, and other cereals"). For the UPD, the original cereals item code (R1-R3 = 108) was recoded to 1083 and presented alongside codes 1081 and 1082 for their respective rounds.

The Section J1 dataset has been divided into 8 categorical datasets to control the size of the individual datasets. The datasets and corresponding food items are as below:

UPD4_HH_J1_1.dta:	"Cereals and Cereal Products"	**hj_00**: 101-112
UPD4_HH_J1_2.dta:	"Starches"	**hj_00:** 201-207
UPD4_HH_J1_3.dta:	"Sugar and Sweets"	**hj_00:** 301-303
UPD4_HH_J1_4.dta:	"Pulses, Dry + Nuts and Seeds"	**hj_00:** 401, 501-504
UPD4_HH_J1_5.dta:	"Vegetables + Fruits"	**hj_00:** 601-603, 701-704
UPD4_HH_J1_6.dta:	"Meat, Meat Products, Fish"	**hj_00:** 801-810
UPD4_HH_J1_7.dta:	"Milk and Milk Products" + "Oil and Fats" + "Spices"	**hj_00:** 901-903, 1001-1004
UPD4_HH_J1_8.dta:	"Beverages"	**hj_00:** 1101-1108

SECTION J3: CONSUMPTION OF FOOD OVER THE PAST ONE WEEK - FREQUENCY OF CONSUMED ITEMS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_J3.dta
Unique Identifier(s):	UPHI, round, hj3_00 **or** UPHI, r_hhid hj3_00	**Rounds Present:**	R2, R3, R4

Description:
Number of days general food categories where consumed by the household in the last 7 days, includes: "Cereals, Grains and Cereal Products", "Roots, Tubers, and Plantains", Nuts and Pulses", Vegetables", Meat, Fish and Animal Products", "Fruits" "Milk/Milk Products", "Fats/Oils", Sugar/Sugar Products/Honey", and "Spices/Condiments".

Original Data Files:

R1:	Not present	**R3:**	HH_SEC_J3.dta
R2:	HH_SEC_K2.dta	**R4:**	HH_SEC_J3.dta

Key Notes:
None

SECTION J4: QUANTITY OF MEALS AND PEOPLE CONSUMING MEALS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_J4.dta
Unique Identifier(s):	UPHI, round, hj4_00 **or** UPHI, r_hhid, hj4_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Number of people living outside of the household that shared meals within the household and then total number of meals shared over the last 7 days.

Original Data Files:

R1: SEC_H1_J_K2_O2_P1_Q1_S1.dta **R3:** HH_SEC_J4.dta
R2: HH_SEC_K3.dta **R4:** HH_SEC_J4.dta

Key Notes:
None

SECTION K: NON-FOOD EXPENDITURES – PAST ONE WEEK & ONE MONTH

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_K.dta
Unique Identifier(s):	UPHI, round, hk_00 **or** UPHI, r_hhid, hk_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Total household expenditure on frequently purchased non-food items during the last week or last month, including public transportation, fuels, cellular phone credits, personal hygiene items, etc.

Original Data Files:

R1: SEC_L.dta **R3:** HH_SEC_K.dta
R2: HH_SEC_L.dta **R4:** HH_SEC_K.dta

Key Notes:
Expenditure on "Repairs and maintenance to dwelling" in the past one month was only collected in R1 and R2.

SECTION L: NON-FOOD EXPENDITURES – PAST TWELVE MONTHS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_L.dta
Unique Identifier(s):	UPHI, round, hl_00 **or** UPHI, r_hhid, hl_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Total expenditure on less frequently purchased non-food items during the past 12 months, including: household items, community contributions, fees and fines, marriage costs, clothing, etc.

Original Data Files:

R1: SEC_L.dta (items 301-317)
SEC_M.dta (all other items)
R2: HH_SEC_M.dta

R2: HH_SEC_L.dta
R4: HH_SEC_L.dta

Key Notes:
For explanation of data derivation only: the R1 Non-Food Expenditures – Past Twelve Months section (Module M) included items 301-317 (with two content questions only) in the R1 Module L dataset (**SEC.L.dta**), as that module had the same two content questions; the remaining items (with three content questions) were in the R1 Module M dataset (**SEC.M.dta**). In the UPD dataset, items were standardized across rounds rather than content questions; thus, the comprehensive list of items in the Non-Food Expenditures – Past Twelve Months Module will be included, and the content questions will be blank for items that are not applicable.

The question wording (and thus the variable label) varies nominally between items 301-324 and items 325-326; the former asks "Over the past 12 months, did you purchase or pay for any […]?" while the latter asks "Over the past 12 months, did you gather, purchase, or pay for any […]? "These have been combined under the UPD variable **hlq01**. Please see the survey instrument for UPD Module L for item specific question wording.

Likewise, these have been combined for the **hlq02** variable, where the question for items 301-324 is "How much did you pay in total?" and "What was the estimated total value of […] consumed?" for items 325-326.

SECTION M: HOUSEHOLD ASSETS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_M.dta
Unique Identifier(s):	UPHI, round, hm_00 **or** UPHI, r_hhid, hm_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Quantity of key items owned by the household, age of item, purchase price, and current estimated value for items including, radio, telephones, appliances, furniture, cookware, vehicles, land, and agricultural tools.

Original Data Files:

R1: SEC_N.dta
R2: HH_SEC_N.dta
R3: HH_SEC_M.dta
R4: HH_SEC_M.dta

Key Notes:
None

SECTION N: FAMILY / HOUSEHOLD NON-FARM ENTERPRISES

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_N.dta
Unique Identifier(s):	UPHI, round, hn_00 **or** UPHI, r_hhid, hn_00	**Rounds Present:**	R3, R4

Description:
Details on non-farm businesses operated by the household during the last 12 months, including type of product or service provided, household members involved with the enterprise, value of current stock, operation location, length of time the enterprise has been operating, income earned, profits, business operating costs etc.

Original Data Files:

R1: Not present
R2: Not present
R3: HH_SEC_N.dta
R4: HH_SEC_N.dta

Key Notes:
Filter questions for UPD Module N are included in the **Filters.dta.** All households are included in the dataset regardless of their filter question to facilitate household-level merges; however, please use the filter question(s) to determine if data are eligible but missing, or if they are missing because there are no non-farm enterprises and responses are thus not applicable.

SECTION O1: ASSISTANCE AND GROUPS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_O1.dta
Unique Identifier(s):	UPHI, round, ho_00 **or** UPHI, r_hhid, ho_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Financial and in-kind assistance provided by both government and non-government institutions, including free food/maize distribution, food for work or cash for work programmes, inputs for work programmes, and school scholarships.

Original Data Files:

R1: SEC_O1.dta
R2: HH_SEC_O1.dta
R3: HH_SEC_O1.dta
R4: HH_SEC_O1.dta

SECTION O2: ASSISTANCE AND GROUPS

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_O2.dta
Unique Identifier(s):	UPHI, round **or** UPHI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
Detailed information on credit and savings groups (SACCOS) that household members are members of, including balances, contributions, and repayment time.

Original Data Files:

R1: SEC_O3.dta
R2: HH_SEC_O2.dta
R3: HH_SEC_O2.dta
R4: HH_SEC_O2.dta

Key Notes:
None

SECTION P: CREDIT

Level of Observation: **Unique Identifier(s):**	Household UPHI, round, hp_00 **or** UPHI, r_hhid, hp_00	**UPD Data File:** **Rounds Present:**	UPD4_HH_P.dta R1, R2, R3, R4

Description:
Utilization of credit and loan institutions, including sources, amounts borrowed, and repayment plans.

Original Data Files:

R1: SEC_P2.dta
R2: HH_SEC_P2.dta
R3: HH_SEC_P.dta
R4: HH_SEC_P.dta

Key Notes:
Filter questions found in **Filters.dta**

SECTION Q1: FINANCE

Level of Observation: **Unique Identifier(s):**	Household UPHI, round **or** UPHI, r_hhid	**UPD Data File:** **Rounds Present:**	UPD4_HH_Q1.dta R2, R3, R4

Description:
Use of financial institutions and mobile-phone based money transfer services, income from rent or pensions, supplemental income sources, and banking history.

Original Data Files:

R1: Not present
R2: HH_SEC_Q.dta
R3: HH_SEC_Q1.dta
R4: HH_SEC_Q1.dta

Key Notes:
None

SECTION Q2: FINANCE

Level of Observation: **Unique Identifier(s):**	Household UPHI, round, sourceid **or** UPHI, r_hhid, sourceid	**UPD Data File:** **Rounds Present:**	UPD4_HH_Q2.dta R2, R3, R4

Description:
Remittances or financial assistance received in the last 12 months from sources residing in Tanzania or abroad, including socio-demographic characteristics of source, remittance channels, and use of cash/in-kind items.

Original Data Files:

R1: Not present
R2: HH_SEC_Q.dta
R3: HH_SEC_Q2.dta
R4: HH_SEC_Q2.dta

Key Notes:
R3 & R4 collected Finance information at the household-level (i.e. one data row; dataset: **HH_SEC_Q1.dta**), and at the source-level (dataset: **HH_SEC_Q2.dta**). While most of the questions align across the three rounds, R2 collected information at household-level only. To enhance universality, R2 has been split into two datasets to mirror R3 & R4.

The resulting dataset **UPD4_HH_Q2.dta** will have no more than one source present for R2 households (based solely on questionnaire formatting and not actual source count), while R3 & R4 households may have more than one source as multiple sources were allowed.

SECTION R: RECENT SHOCKS TO HOUSEHOLD

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_R.dta
Unique Identifier(s):	UPHI, round, hr_00 **or** UPHI, r_hhid, hr_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Severity ranking of 16 common shocks, such as "drought or floods", "severe water shortage", and "death of member of household". More detailed information and coping strategies are asked of the three most severe shocks.

Original Data Files:

R1: SEC_R.dta
R2: HH_SEC_R.dta
R3: HH_SEC_R.dta
R4: HH_SEC_R.dta

Key Notes:
The original R1 Module R dataset only provided data for the specific shock(s) affecting each respective household – i.e. if the answer was "No" to a shock in the shock listing, no data line was recorded (rather than the comprehensive listing present for each household in the R2, R3, & R4 datasets and the first question recording "Yes"/"No".

Therefore, households in R1 will have any number of data lines between 0-12 (19 is the total possible, but the most shocks any household reported was 12).

For this module (and others like it) it is important to consider the *possible* denominator when calculating rates. For example, **hrq07_1** had 20 possible response codes in R3, but only 15 in R4. While they maintain the same numeric response code value across years, a tabulation will report "0" for the 5 response codes not presented in R4. These are not truly "0", rather were not possible in the round-specific questionnaire; thus, the UPD survey instrument should be consulted prior to all analyses.

SECTION S: DEATHS IN HOUSEHOLD

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_S.dta
Unique Identifier(s):	UPHI, round, hs_00 **or** UPHI, r_hhid, hs_00	**Rounds Present:**	R1, R2, R3, R4

Description:
Deaths in the household within the past two years, cause of death, duration of illness if applicable, and associated land or asset losses due to inheritance traditions.

Original Data Files:

R1: SEC_S2.dta
R2: HH_SEC_S.dta
R3: HH_SEC_S.dta
R4: HH_SEC_S.dta

Key Notes:
None

SECTION V: ANTHROMPOMETRY

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_V.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid	**Rounds Present:**	R1, R2, R3, R4

Description:
Anthropometric information was collected from all household members who were able and willing to participate in R1-R3, and for those under the age of 15 and women of child bearing age (15-49 years old) in R4, including height, weight, and upper arm circumference measurements.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta
R2: HH_SEC_U.dta
R3: HH_SEC_V.dta
R4: HH_SEC_V.dta

Key Notes:
None

SECTION W1: CHILDREN LIVING ELSEWHERE

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_W1.dta
Unique Identifier(s):	UPI, round **or** UPI, r_hhid **or** round, r_hhid, r_id	**Rounds Present:**	R1

Description:
Two filter questions for the "Children Living Elsewhere" module which requests information on female household members age 27 and older, and their children age 15 and older living elsewhere.

Original Data Files:

R1: SEC_B_C_D_E1_F_G1_U.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
None

SECTION W2: CHILDREN LIVING ELSEWHERE

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_W2.dta
Unique Identifier(s):	UPI, child_id	**Rounds Present:**	R1

Description:
Details on individuals (>15 years of age) living outside of the NPS household, including brief information on education, occupation, and remittances. Information in this module is completed by the mother of the individual and not by the individual who is not in the household.

Original Data Files:

R1: SEC_G2.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
This module only contains information for women greater >27 years of age who have children >15 years of age living outside of the household. While individual-level module merges with **UPD4_HH_W1.dta** will result in perfect merges, **UPD4_HH_W2.dta** is a supplementary dataset resulting from the filter questions in Module W1.

SECTION X1: GOVERNANCE

Level of Observation: **Unique Identifier(s):**	Household UPHI, round **or** UPHI, r_hhid	**UPD Data File:** **Rounds Present:**	UPD4_HH_X1.dta R1, R2

Description:
This section collects general information on the primary respondent for the governance section.

Original Data Files:

R1: SEC_H1_J_K2_O2_P1_Q1_S1.dta
R2: HH_SEC_H3.dta
HH_SEC_FILTERS.dta
R3: Not present
R4: Not present

Key Notes:
None

SECTION X2: GOVERNANCE

Level of Observation: **Unique Identifier(s):**	Household UPI, round, hx2_00 **or** UPI, r_hhid, hx2_00	**UPD Data File:** **Rounds Present:**	UPD4_HH_X2.dta R1, R2

Description: Details on knowledge of government officials (if known by respondent), including approval.

Original Data Files:

R1: SEC_H2.dta
R2: HH_SEC_H1.dta
R3: Not present
R4: Not present

Key Notes:
None

SECTION X3: GOVERNANCE

Level of Observation: **Unique Identifier(s):**	Household UPI, round, hx3_00 **or** UPI, r_hhid, hx3_00	**UPD Data File:** **Rounds Present:**	UPD4_HH_X3.dta R1, R2

Description: Information on the primary respondent's attendance at local governance meetings, religious affiliation, and participation in religious activities.

Original Data Files:

R1: SEC_H3.dta
R2: HH_SEC_H2.dta
R3: Not present
R4: Not present

Key Notes:
None

SECTION Y: VIOLENCE AGAINST WOMEN

Level of Observation:	Individual	**UPD Data File:**	UPD4_HH_Y.dta
Unique Identifier(s):	UPHI, indidy1	**Rounds Present:**	R1

Description: Attitudes about violence against women and the types of violence experienced.

Original Data Files:

R1: SEC_I.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
This module was only asked of women ages 15-50.

SECTION Z1: CRIME & JUSTICE

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_Z1.dta
Unique Identifier(s):	UPHI	**Rounds Present:**	R1

Description:
Questions to determine if any crimes were suffered by the household during the twelve months prior to the survey (Section Z2) or if anyone in the household was involved in disputes (Section Z3).

Original Data Files:

R1: SEC_H1_J_K2_O2_P1_Q1_S.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
To properly allow for merges back to other UPD datasets, the UPHI variable has been included and thus, there are duplicates at the **y1_hhid** level (because of the various panel combinations using a single R1 household). The UPHI variable will be the unique identifier for the module but analyses will most likely occur at the **y1_hhid** level.

SECTION Z2: CRIME & JUSTICE

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_Z2.dta
Unique Identifier(s):	UPHI, hz_00	**Rounds Present:**	R1

Description:
Information on crimes suffered by the household during the last twelve months (Section Z2) and associated interactions with the police and justice system (see Section Z3).

Original Data Files:

R1: SEC_Q2.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
None

SECTION Z3: CRIME & JUSTICE

Level of Observation:	Household	**UPD Data File:**	UPD4_HH_Z3.dta
Unique Identifier(s):	UPHI, hz_00	**Rounds Present:**	R1

Description:
Information on disputes suffered by the household during the last twelve months (see Section Z2) and associated interactions with the police and justice system (Section Z3).

Original Data Files:

R1: SEC_Q3.dta
R2: Not present
R3: Not present
R4: Not present

Key Notes:
None

Appendix

Appendix I: NPS UPD Naming Convention

R1		R2		R3		R4		UPD	
A	Household Identification & Survey Staff Details	**A**	Household Identification & Survey Staff Details	**A**	Household Identification & Survey Staff Details	**A**	Household Identification & Survey Staff Details	**A**	Household Identification & Survey Staff Details
B	Household Member Roster	**B**	Household Member Roster	**B**	Household Member Roster	**B**	Household Member Roster	**B**	Household Member Roster
C	Education	**C**	Education	**C**	Education	**C**	Education	**C**	Education
D	Health	**D**	Health	**D**	Health	**D**	Health	**D**	Health
E	Labor	**E**	Labor	**E**	Labor	**E**	Labor	**E**	Labor
F	Food Outside the Household	**F**	Food Outside the Household	**F**	Food Consumption Outside the Household	**F**	Food Consumption Outside the Household	**F**	Food Consumption Outside the Household
		G	Subjective Welfare	**G**	Subjective Welfare	**G**	Subjective Welfare & Crime	**G**	Subjective Welfare & Crime
		I	Food Security	**H**	Food Security	**H**	Food Security	**H**	Food Security
J	Housing, Water & Sanitation	**J**	Housing, Water & Sanitation	**I**	Housing, Water & Sanitation	**I**	Housing, Water & Sanitation	**I**	Housing, Water & Sanitation
K	Consumption of Food Over Past One Week	**K**	Consumption of Food Over Past One Week	**J**	Consumption of Food Over Past One Week	**J**	Consumption of Food Over Past One Week	**J**	Consumption of Food Over Past One Week
L	Non-Food Expenditures	**L**	Non-Food Expenditures	**K**	Non-Food Expenditures	**K**	Non-Food Expenditures	**K**	Non-Food Expenditures
M	Non-Food Expenditures	**M**	Non-Food Expenditures	**L**	Non-Food Expenditures	**L**	Non-Food Expenditures	**L**	Non-Food Expenditures
N	Household Assets	**N**	Household Assets	**M**	Household Assets	**M**	Household Assets	**M**	Household Assets
				N	Family/Household Non-Farm Enterprises	**N**	Family/Household Non-Farm Enterprises	**N**	Family/Household Non-Farm Enterprises
O	Assistance and Groups	**O**	Assistance and Groups	**O**	Assistance and Groups	**O**	Assistance and Groups	**O**	Assistance and Groups
P	Credit	**P**	Credit	**P**	Credit	**P**	Credit	**P**	Credit
		Q	Finance	**Q**	Finance	**Q**	Finance	**Q**	Finance
R	Recent Shocks to Household Welfare	**R**	Recent Shocks to Household Welfare	**R**	Recent Shocks to Household Welfare	**R**	Recent Shocks to Household Welfare	**R**	Recent Shocks to Household Welfare
S	Deaths in Household	**S**	Deaths in Household	**S**	Deaths in Household	**S**	Deaths in Household	**S**	Deaths in Household
V	Household Re-Contact Information	**V**	Household Re-Contact Information	**U**	Household Re-Contact Information	**U**	Household Re-Contact Information	**U**	Household Re-Contact Information
U	Anthropometry	**U**	Anthropometry	**V**	Anthropometry	**V**	Anthropometry	**V**	Anthropometry
G	Children Living Elsewhere							**W**	Children Living Elsewhere
H	Governance	**H**	Governance					**X**	Governance
I	Violence Against Women							**Y**	Violence Against Women
Q	Crime & Justice							**Z**	Crime & Justice

Appendix II: Questionnaire Footnotes

MODULE B

[1] R1 question: "IF THIS IS AN HBS HOUSEHOLD MEMBER, ENTER THE HBS ID NUMBER"

[2] R1 lists "UNPAID FAMILY WORK" twice; values/labels in R1 dataset indicate code 12 is "PAID FAMILY WORK" as in other rounds.

[3] R1 specified "ENTER 0 IF NONE"; removed from UPD.

[4] "DISTRICT/COUNTRY NAME" and "REGION CODE" are included on R1 questionnaire, but variables are not included in R1 datasets so no R1 variable name is presented here.

MODULE C

[1] The UPD question has been altered to be applicable to all rounds. See round-specific questionnaires for exact wording.

[2] "ONE WAY ONLY" only specified in R3 & R4 questionnaires.

MODULE D

[1] In R1, respondents 12 years and older were instructed to answer for themselves, so R1 questions are worded with respect to the individual answering the question, i.e. "Have you visited a health care provider…". "[NAME]" was used in R2, R3, and R4 and thus is the format used in the text of the UPD question. Only the first instance is noted but this applies to all questions in the R1 questionnaire.

[2] R2 question: "During the past 12 months, what measures were taken to improve [NAME]'s performance of activities?"

[3] "SPECIFY" was added to response code "OTHER" beginning in R2. The numeric response code (=7) remains the same across rounds.

[4] R1 & R2 question: "Was this birth registered?"

MODULE E

[1] In R2, the questions were worded in respect to the individual answering the question, i.e. "Did you do any work of any type…". "[NAME]" was used in R1, R3, and R4 and thus is the format used in the text of the universal panel question. Only the first instance is noted but this applies to the remainder of questions in the R2 questionnaire.

[2] The UPD question has been altered to be applicable to both R3 & R4. See round-specific questionnaires for exact wording.

[3] The instruction "(WAGE EMPLOYEMENT)" is only specified in the R1 questionnaire; however, in each of R1 and R2 this question is included in the "Wage Jobs" section.

[4] See round-specific questionnaires for slight variations in wording; differences not expected to alter interpretation.

[5] The instruction "(if any)" is only specified in the R1 questionnaire, while the instruction "[ENTER '0' / '0' IF NEVER]" is only in R2.

[6] The instruction "[CODE: TASCO CODE]" is only present on the R3 & R4 questionnaires.

[7] The instruction "[APART FROM SALARY]" is only specified in the R1, R3, & R4 questionnaires.

[8] Maximum amounts of months, weeks, and hours are only specified on the R3 & R4 questionnaires.

[9] Both "[wage]" and "[paid]" are specified in the R3 question, but are only in the section header in the R4 questionnaire.

[10] The R2 & R3 questionnaires specify "wage" in the question, and in the section header in the R4 questionnaire.

[11] Questions in the self-employment section of the R1 questionnaire used the pronoun "you"/"your" instead of "[NAME]", similar to the language used throughout the R2 questionnaire. The text of UPD questions with this footnote is identical to the text in the questions in both R1 and R2 and is meant only to highlight the deviation from the language used throughout the remainder of the R1 questionnaire.

[12] The instruction "IF NONE, WRITE '0'" is only in the R2 questionnaire.

[13] The instruction "[IF NONE, WRITE '0']" is only specified in R1 & R2.

[14] Maximum amount of hours is only specified in R3 & R4.

[15] Note: There is a question number 49 in Module E of R1. This question "seq49_1"-"seq49_6" will be found in UPD Module G (as UPD question hg_03_1-hg_03_9) as they align with the Subjective Welfare & Crime question from R2, R3, & R4.

MODULE F

No footnotes

MODULE G

[1] Response (8) was "DOES NOT APPLY" in R1 and "NOT APPLICABLE" in R2, R3, and R4; the UPD uses "NOT APPLICABLE"

MODULE H

No footnotes

MODULE I1

[1] While both R1 and R2 report two sources, the instruction "UP TO 2 SOURCES" was only specified on the R2 questionnaire.

[2] "SPECIFY" was added to the "OTHER" response code starting in R2. The numeric response code value (=6) remains the same across rounds.

[3] R1 question: "What is the main toilet facilities for this household?" & R2 question: "What is the main toilet facilities usually used in this household?"

[4] "SPECIFY" was added to the "OTHER" response code starting in R2.

[5] R3: "How many other households share this toilet facility?"

[6] SPECIFY" was added to the "OTHER" response code starting in R2. The numeric response code value (=8) remains the same.

[7] "SPECIFY" was added to the "OTHER" response code starting in R2. The numeric response code value (=7) remains the same.

[8] Referred to as "RAINWATER" in R1-R3.

[9] "SPECIFY" was added to the "OTHER" response code starting in R2. Numeric response code values changed over time.

[10] R2 & R3 Qx: "How long does it take to get water from drinking water source to this dwelling in the rainy season?";

[11] R2 & R3 Qx: "Out of these [READ] minutes, how long do you spend waiting?"

[12] R2 & R3 Qx: "How long does it take to get water from drinking water source to this dwelling in the dry season?";

[13] R2 & R3 Qx: "Out of these [READ] minutes, how long do you spend waiting?"

[14] "BOIL" (R2 & R3) was referred to as "BOILING WATER" in R1. The numeric response code value remains the same; 15"MARK UP TO 2" was only specified in R2 & R3.

[15] "MARK UP TO 2" was only specified in R2 & R3.

MODULE J1

No footnotes

MODULE J3

No footnotes

MODULE J4

No footnotes

MODULE K

[1] "Mosque" was added to code 218 in R3 and R4

MODULE L

No footnotes

MODULE M

No footnotes

MODULE N

No footnotes

MODULE O1

No footnotes

MODULE O2
[1] The letter ID was not included in the R1 datasets and thus does not have a variable name like the other rounds.
MODULE P
[1] "in the last 12 months" was not included in the R2 or R1 question. *[2] The instruction "LIST UP TO THREE IN ORDER OF IMPORTANCE" was not specified in R2.*
MODULE Q1
[1] R2 question: Which of the following remittances channels did you use in the last 12 months?
MODULE Q2
[1]R2 question: What is your relationship to the sender? *[2]R2 question: "From what country did these remittances originate?* *[3]R2 question: Which of the following remittances channels did you use in the last 12 months?* *[4] R2 question: When you receive the money from abroad what did you use it for?*
MODULE R
[1] Updated from "2" to "UP TO 3" to reflect maximum number allowed by rounds (=3), R3.
MODULE S
[1] The phrase "including those listed as "dead" in PRE-PRINTED TRACKING FORM" was added in R2.
MODULE V
No footnotes
MODULE W1
No footnotes
MODULE W2
[1]Some code values and value labels have been altered slightly between the R1 and UPD versions (see "NEVER ATTENDED", F1-F4, and U2-U3). *[2]R1 codes differ between R1 questionnaire and R1 dataset; UPD has made the best effort to standardize and correct these.*
MODULE X1
[1] The instruction "FROM THE PRE-PRINTED TRACKING FORM" was only specified in the R2 question.
MODULE X2
[2] The instruction "IF POST IS VACANT, WRITE '00'" was only specified in the R2 question.
MODULE X3
No footnotes
MODULE Y
No footnotes
MODULE Z1
No footnotes
MODULE Z2
[1] UPD question wording changed from "above" to "in previous questions".
MODULE Z3
[2] UPD question wording changed from "above" to "in previous questions".