



Data Description

RWI – Leibniz Institute for Economic Research

**FDZ Data Description:
The Senegal Migration Panel (SMP)
User Guide 2 – Wave 1 (2022)**

Cara Ebert, Bernd Beber, Zara Riaz

March 2026



Imprint

Editor

RWI – Leibniz Institute for Economic Research e. V.
Hohenzollernstraße 1–3 | 45128 Essen, Germany

P.O. Box: 10 30 54 | 45030 Essen, Germany
Fon: +49 201-81 49-0 | E-Mail: rwi@rwi-essen.de
www.rwi-essen.de

Board of Directors

Prof. Dr. Dr. h.c. Christoph M. Schmidt (President)
Prof. Dr. Thomas K. Bauer (Vice President)
Dr. Stefan Rumpf (Administrative Board Member)
Prof. Dr. Kerstin Schneider (Member of the extended Executive Board)

© RWI 2026

Reprinting, even in extracts, is only permitted with the permission of the RWI.

RWI Data Description

Editor: Prof. Dr. Dr. h. c. Christoph M. Schmidt
Layout: Magdalena Franke, Claudia Lohkamp

FDZ Data Description:
The Senegal Migration Project (SMP)
User Guide 2 - Wave (2022)

March 2026

Project team

RWI: Cara Ebert, Bernd Beber, Zara Riaz



The RWI is funded by the federal government
and the state of North Rhine-Westphalia.

Inhaltsverzeichnis

1.	Introduction	5
2.	Description of surveys	5
3.	Data access	6
3.1	Data access	6
3.2	Anonymization.....	6
3.3	Variables subject to embargo.....	7
4.	Sampling	8
5.	Survey procedures	8
5.1	Survey software and forms.....	8
5.2	Respondent selection of the Census, Network and Core-household surveys.....	8
5.3	Respondent selection of the Core-individual survey.....	9
5.4	Selection into/out of Core-household and Core-individual surveys	9
5.5	Fusion of households across data waves.....	10
6.	Using the data.....	12
6.1	Data and materials.....	12
6.2	Unique identifiers	13
6.3	Identifying and linking individual information and respondents	13
6.4	Variable labels.....	15
7.	Recoding of data	15
7.1	Updating the member roster in the <i>Census</i> survey using subsequent data	15
7.2	Household head recodes in the <i>Census</i> survey	16
7.3	Migration roster recodes in the <i>Census</i> survey	16
7.4	Occupation recodes in the <i>Census</i> and <i>Core-individual</i> surveys	16
8.	Animal recodes in the <i>Core-household</i> survey	18

Tabellenverzeichnis

Table 1: 2022 SMP surveys, their respondents, sample sizes and dates.....	5
Table 2: Variables in the Core-household and Core-individual surveys that are currently under embargo	7
Table 3: Household IDs (hhid_anon) of the 2022 phase 2 surveys that fused with another household since phase 1 surveys, and were interviewed as part of that other, now larger, household in phase 2	11
Table 4: Updated list with response codes for occupations in the Census and Core-individual surveys	17

1. Introduction

This **User Guide** provides a description of the **SMP Wave 1 (2022)** surveys:

1. 2022 **Village chief** survey,
2. 2022 **Census**,
3. 2022 **Network** survey,
4. 2022 **Core-household** survey,
5. 2022 **Core-individual** survey.

The User Guide includes information on respondent selection, data collection procedures, and how to use the data.

The 2022 SMP surveys were funded by the *Policy Lab “Climate Change, Development and Migration”*¹ of the RWI – Leibniz Institute for Economic Research.

A **general introduction to the Senegal Migration Panel (SMP)** can be found in “*SMP - User guide - 1 Overview – 20260330.docx*”, providing an overview of all SMP surveys and waves, information about the sampling of wave 1 (this wave) which all further sampling is based on, and linking households and individuals across surveys and waves.

2. Description of surveys

Table 1 presents an overview of the different 2022 SMP surveys, their respondents, their sample size and timing. In a first survey phase the *Village chief* and *Census* surveys were conducted. In a second survey phase the *Network*, *Core-household* and *Core-individual* surveys were conducted.

Table 1: 2022 SMP surveys, their respondents, sample sizes and dates.²

	Respondent	N sampled	N consented	Dates
Village chief	Village chief	145	144	Jul. 16, 2022 - Aug. 14, 2022
Census	HH head or knowledgeable person in HH	9,937	9,665	Jul. 16, 2022 - Aug. 12, 2022
Network	HH head or knowledgeable person in HH + only for network questions: the respondent of the <i>Core-individual</i> survey	9,937	9,359	Aug. 27, 2022 - Oct. 22, 2022
Core-household (in-person)	HH head or knowledgeable person in HH	6,031	4,057	Aug. 27, 2022 - Oct. 22, 2022
Core-household (phone)	Male 18 to 40 years		751	Oct. 7, 2022 - Oct. 24, 2022
Core-individual (in-person)	Male 18 to 40 years	6,031	3,935	Aug. 27, 2022 - Oct. 22, 2022
Core-individual (phone)			751	Oct. 7, 2022 - Oct. 24, 2022

Source: Own Data.

¹ More information on the Policy Lab “Climate Change, Development and Migration” can be found here: [Policy Lab Climate Change, Development, Migration - RWI Essen](#).

² To track sample sizes of the *Core-household* and *Core-individual* surveys cross-tab the variables *respondent_status* and *int_type*.

First survey phase:

The first survey phase took place between mid-July and mid-August 2022.

In each village, the *Village chief* survey was conducted first to gain access to the community, and subsequently enumerators proceeded with the interviews for the *Census* survey.

In all 145 SMP villages but one *Village chief* interviews were completed (**N=144, in-person**). The survey included questions about village characteristics and the chief's perceptions about environmental changes.

The *Census* includes **9,665 in-person interviews**, listing all household members of age 15 and older ("HH roster"), information about their basic sociodemographic background and their migration history. The respondent of the *Census* survey was the household head or a person who knows about household matters well.

Second survey phase:

In the second survey, each household included in the *Census* survey of the first phase, was followed up in-person between end of August and end of October 2022.

First, a household survey was conducted with the household head or a member who knows about household matters well.

The household was asked a short set of socioeconomic background questions and network questions (**Network survey, N= 9,359 in-person**). For the network module the respondent of the *Core-individual* survey (see below) was asked to join the interview to ensure that the household network also represents the individual's connections.

If the household had been selected for the *Core-household and -individual* surveys, the household was additionally asked an extended list of more detailed socioeconomic information about the household and climate change perceptions of the respondent (**Core-household survey, N=4,057 in-person**).

Second, in households selected for the *Core-household* and *Core-individual* surveys an individual survey was conducted with a male aged 18 to 40 years about individual socioeconomic characteristics, migration behavior and intentions, wellbeing and risk preferences among other things (**Core-individual survey, N=3,935 in-person**).

Third, when the *Core-household and Core-individual* interviews could not be conducted in-person and only the *Network* survey was completed, **phone mop-up** interviews were conducted with the respondent of the *Core-individual* survey (**N=751**). In addition to the questions from the *Core-individual* survey, the respondent was also asked a short set of questions from the *Core-household* survey.

3. Data access

3.1 Data access

The datasets are available free of charge for scientific purposes as a Scientific Use File (SUF) Off-site at <https://doi.org/10.7807/SMP:2022:V1>.

3.2 Anonymization

Identifying information such as names, phone numbers, GPS data and village names were removed from the dataset. The variable *ea_anon* includes anonymized village identifiers that were assigned within regions and departments.

3.3 Variables subject to embargo

Embargo restrictions exist for the *Core-household* and *Core-individual* surveys. Table 2 lists the variables that are currently subject to a release embargo. In the present version of the dataset, their values have been replaced with the placeholder “EMBARGO”. The embargo is expected to be lifted in 2028, after which the variables will be released in full as part of an updated version of the dataset under the same DOI.

Table 2: Variables in the *Core-household* and *Core-individual* surveys that are currently under embargo

hh_env_change_1	networks_q1_other_gender_*	networks_q4_roster_count
hh_env_change_2	networks_q1_other_hfirst_*	networks_q4_index_*
hh_env_change_3	networks_q1_other_hlast_*	networks_q4_entry_*
hh_env_change_4	networks_q1a_family_*	networks_q4_select_anon*
hh_env_change_5	networks_q1a_conc_*	networks_q4_other_nfirst_*
hh_rainfall_quality_2021	networks_q2_num	networks_q4_other_nlast_*
hh_rainfall_problem_2021	networks_q2_roster_count	networks_q4_other_age_*
hh_rainfall_problem_2021_other	networks_q2_index_*	networks_q4_other_gender_*
hh_rainfall_quality_2022	networks_q2_entry_*	networks_q4_other_hfirst_*
rainfall_problem_2022	networks_q2_select_anon*	networks_q4_other_hlast_*
hh_rainfall_problem_2022_other	networks_q2_other_nfirst_*	networks_q4a_family_*
hh_rainfall_quality_2023	networks_q2_other_nlast_*	networks_q4a_conc_*
hh_shock	networks_q2_other_age_*	networks_q5_num
hh_shock_other	networks_q2_other_gender_*	networks_q5_roster_count
hh_shock_roster_count	networks_q2_other_hfirst_*	networks_q5_index_*
hh_shock_id_*	networks_q2_other_hlast_*	networks_q5_entry_*
hh_shock_name_*	networks_q2a_family_*	networks_q5_select_anon*
hh_shock_year_*	networks_q2a_conc_*	networks_q5_other_nfirst_*
hh_shock_affect_inc	networks_q3_num	networks_q5_other_nlast_*
hh_disasters_frequency	networks_q3_roster_count	networks_q5_other_age_*
hh_shock_perception	networks_q3_index_*	networks_q5_other_gender_*
hh_cc_adaptation_1	networks_q3_entry_*	networks_q5_other_hfirst_*
hh_cc_adaptation_2	networks_q3_select_anon*	networks_q5_other_hlast_*
hh_cc_adaptation_3	networks_q3_other_nfirst_*	networks_q5a_family_*
networks_q1_num	networks_q3_other_nlast_*	networks_q5a_conc_*
networks_q1_roster_count	networks_q3_other_age_*	i_env_respondent
networks_q1_index_*	networks_q3_other_gender_*	i_env_change_1
networks_q1_entry_*	networks_q3_other_hfirst_*	i_env_change_2
networks_q1_select_anon*	networks_q3_other_hlast_*	i_env_change_3
networks_q1_other_nfirst_*	networks_q3a_family_*	i_env_change_4
networks_q1_other_nlast_*	networks_q3a_conc_*	i_env_change_5
networks_q1_other_age_*	networks_q4_num	

Source: Own data.

4. Sampling

The sampling for all surveys (Village chief, Network, Core-household and Core-individual) is described in “*SMP - User guide - 1 Overview – 20260330.docx*”.

See Section 5 “Survey procedures” of this User Guide for a description of how cases were handled when respondents were not available.

5. Survey procedures

5.1 Survey software and forms

Data collection was conducted using SurveyCTO Collect on Tablets.

The in-person interviews of the *Network*, *Core-household* and *Core-individual* surveys were conducted in one questionnaire form to allow for a consistent sample and respondent selection. The *Village chief*, *Census* and *Core* phone mop-up surveys were using separate forms.

5.2 Respondent selection of the Census, Network and Core-household surveys

The respondent of the *Census*, *Network* and *Core-household* surveys was the head of household or a member who knows about household matters well.

Note that we did not require the respondent of the *Network* or *Core-household* surveys to be the same person as the respondent of the *Census* survey in the first phase. The respondent of the *Network* and *Core-household* surveys were usually the same as they were conducted during the same household visit.

When no head or knowledgeable member was available, but the respondent of the *Core-individual* survey was available, the *Network* and *Core-household* interviews proceeded with the respondent of the *Core-individual* interview.

Variable *respondent_status* indicates the respondent status of the interview.

If *respondent_status* equals “available”, all required respondents were available. In cases where only a Network Survey was assigned the only required respondent was the household head or a knowledgeable member. In cases where the *Network*, *Core-household*, and *Core-individual* surveys were assigned, this was the household head or a knowledgeable member, and the individual respondent.

If *respondent_status* equals “r2_na”, the household head or a knowledgeable member were available for the interview, but not the respondent of the *Core-individual* interview.

If *respondent_status* equals “r1_na”, the household head or a knowledgeable member were *not* available for the interview and the household interview was conducted with the respondent of the *Core-individual* survey. Note that often the respondent of the *Core-individual* survey was knowledgeable about household matters and therefore responded to the household interview, but in these cases *respondent_status* would not indicate “r1_na” (but “available” instead).

If *respondent_status* equals “r1_r2_na”, no interview was conducted as neither the household head nor a knowledgeable member were available for the *Network* and *Core-household* interviews, and also no respondent for the *Core-individual* interview was available.

Section 6.3. “Identifying and linking individual information and respondents” lists the variables that include the IDs of household members responding to the respective survey module and describes how to link individuals across surveys.

5.3 Respondent selection of the Core-individual survey

Eligible respondents of the *Core-individual* survey and their replacements were sampled prior to data collection based on data from the *Census* survey. If the sampled respondent was not available at the time of the data collection household visit, the next male on a list (with all males 18 to 40 years in the household ordered randomly) was selected for an interview. When none of those previously sampled were available, the household was asked if there is a male aged 18 to 40 years that had not been recorded in the *Census* and that could be surveyed. Variables *respondent2** in the questionnaire reflect the procedure of the respondent selection process but were dropped from the final dataset for clarity and better use. The variable *target_position* indicates the position of the selected individual in the list of eligible respondents.

If the *Core-individual* interview could not be conducted in-person even after multiple attempts, a phone mop-up survey was conducted following the same respondent selection procedure.

Variable *ind_id_anon* indicates the ID of the household member responding to the *Core-individual* survey. This individual responds to all SMP *Core-individual* surveys across survey waves.

5.4 Selection into/out of Core-household and Core-individual surveys

The *Core-household* and *Core-individual* surveys were **only conducted when there was an eligible respondent for the *Core-individual* survey** available (see above how these were selected). If no eligible respondent was available, the household only responded to the *Network* survey.

However, as an exception to this, there are 122 households for which a *Core-household* interview was conducted but no *Core-individual* interview. This is because at the start of the interview it was indicated that an eligible respondent would be available for the individual component, but the individual respondent was not actually present/available at the time of the individual interview after the household and network components were completed.

Variable *int_type* indicates the **sampling assignment** of consented households to the *Network* survey and *Core-household* and *-individual* surveys, **and their actual completion status**. All interviews for which respondent consent is obtained, either in the in-person or phone mop-up interview, have a non-missing value for *int_type*.

- *int_type* = 1: households that were **assigned** to the *Network* survey and therefore only completed a short module on household-level assets and network questions.
- *int_type* = 2: households that were **assigned** to the *Core-household* and *-individual* surveys but interviews **turned into** the shorter *Network* survey because there was no eligible respondent (a male HH member between 18-40 years old) present to complete the *Core-individual* interview.
- *int_type* = 3: households that were **assigned** to the *Core-household* and *-individual* surveys and in which an eligible target was present and thus the **surveys proceeded as assigned**.
 - **Note:** This category includes the 122 households for which a full *Core-household* interview was conducted but no *Core-individual* interview.
 - **Note:** This category also includes 5 households that did not consent to the *Network* or *Core-household* interview in-person, but *Core-household* and *Core-individual* interviews were conducted over the phone: i.e. for these 5 households *phone_consent* equals 1. Note that the phone survey only included a shortened set of household items.

- **Note:** If a phone mop-up survey was successfully conducted (N=751), the interview is placed into *int_type*=3. For these 751 cases, however, a subset of *Core-household* questions was asked to the *Core-individual* respondent (*Network* survey items were already completed in-person), including the following variables:
 - *hh_fin_know*
 - *hh_inc_agric*
 - *hh_inc_nonagric*
 - *hh_inc_other*
 - *crop_sold_quantity*
 - *crop_income_other*
 - *livestock_income*
 - *hh_savings*
 - *hh_savings_type*
 - *hh_savings_type_other*
 - *hunger_worried*
 - *hunger_worried_month*
 - *hunger_actual*
 - *hunger_actual_month*
 - *remittances*
 - *remittances_months*
 - *remittances_frequency*
 - *remittances_amount*
 - *hh_cc_adapation_3*

The variable **status** indicates the survey participation status of the household irrespective of the interview type indicated by *int_type* and states the reason if a household was not surveyed.

- *status = Completed:* Interview was completed.
- *status = Refused:* Household refused participation.
- *status = Not available:* Household was busy and was not surveyed.
- *status = Moved permanently:* Household moved permanently outside the village.³
- *status = Fusion:* Fused with a different household and was therefore not surveyed.
- *status = Can't locate:* Location of household is unknown.
- *status = Incorrectly sampled, HH not part of village:* Household was incorrectly assigned to the village during the Census survey.
- *status = Other:* The reason for non-participation does not fit into the categories above.⁴

5.5 Fusion of households across data waves

Some households fused with other households from the first contact (*Census* survey of phase 1) to the second contact (*Network*, *Core-household* and *Core-individual* surveys of phase 2). Therefore, only one household was enumerated during the second contact. Table 3 lists the original household IDs *hhid_anon* that were interviewed as part of the earlier but not the later survey as

³ Certain values of „*hh_access_other*“ include text indicating that the house is vacant while *moved_permanently*==0. Households are coded as having moved permanently only if this is explicitly indicated in *hh_access_other*, another variable about general interview comments, or the *moved_permanently* variable.

⁴ This category includes miscellaneous responses such as indications that the household is a religious school or dorm, or that a physical dwelling is present but the habitants have not yet moved.

well as the household IDs *fusion_hhid_anon* that the household fused with and an interview was completed during phase 2.

Column “*hhid_anon*” includes the ID of the household that refused the surveys of phase 2 (*Network*, *Core-household* and *Core-individual* surveys) as they had been surveyed already as part of another household then. Column “*fusion_hhid_anon*” specifies the ID of the household the observed household fused with and was surveyed as. If “*fusion_hhid_anon*” is “NA”, then it is known that the phase 2 interviews were refused because they were interviewed already but their fusion household is unknown. If there was a true duplication of a household at the time of the *Census* (i.e. the exact same household was interviewed twice), then columns “*hhid_kept*” and “*hhid_deleted*” indicate which observation was kept and which one deleted.

In the dataset, the variable *fusion_hh* specifies the ID of the household the observed household fused with. The variable *fusion_hh* has a value regardless of whether it is the master household (status = “Completed”) or the household that fused with it (status = “Fusion”).

Table 3: Household IDs (*hhid_anon*) of the 2022 phase 2 surveys that fused with another household since phase 1 surveys, and were interviewed as part of that other, now larger, household in phase 2

<i>hhid_anon</i>	<i>fusion_hhid_anon</i>	<i>hhid_kept</i>	<i>hhid_deleted</i>
01003-007	01003-005	01003-007	01003-005
00812-043	00812-042		
00401-035	00401-026		
00410-005	00410-004		
00420-036	00420-035		
00420-072	00420-046		
01027-004	01027-012		
00805-042	00805-050		
00805-065	00805-051		
00806-072	00806-070		
00809-049	00809-018		
00812-007	00812-009		
00816-027	00816-026		
00816-035	00816-036		
00816-042	00816-031		
00816-046	00816-045		
00816-050	00816-049		
00816-051	00816-052		
00816-054	00816-055		
00816-059	00816-058		
00818-073	00818-131		
00818-141	00818-126		
00828-069	00828-067		
00828-070	00828-067		
00829-014	00829-003		
00830-067	00830-066		

Table 3 continued

hhid_anon	fusion_hhid_anon	hhid_kept	hhid_deleted
00703-014	NA		
00703-015	00703-013		
00703-016	00703-013		
00703-017	NA		
00828-076	NA		
00818-091	NA		
00431-072	NA		
00413-081	NA		
00413-082	NA		
00828-066	NA		
00828-068	NA		
00811-005	NA		
00826-025	NA		
00408-019	00408-023		
00804-022	NA		
00804-001	NA		
00829-010	00829-009		
01004-048	01004-024		
00830-051	00830-057		
00426-004	NA		
00703-005	00703-004		
00303-032	NA		
00703-052	NA		
01004-029	01004-015		
00422-059	00422-071* (members overlap)		
00602-067	00602-016		
00809-073	00809-066		
00431-045	NA		

Source: Own data.

6. Using the data

6.1 Data and materials

The data of the *Network*, *Core-household*, *Core-individual* and phone mop-up surveys are contained in one dataset. The phone mop-up data was merged into the in-person data. The individual in-person and phone surveys have the same variable names when survey items were identical. When variables were adjusted for the phone interview and therefore differed (slightly) or when an in-person interview item was repeated during the phone survey (e.g. consent or enumerator codes), additional variables are included and flagged by the **suffix_phone**. For example, *enum_name_phone* indicates the enumerator code of the phone interview, whereas *enum_name* (without any suffix) refers to the enumerator code of the in-person survey.

Variable *consented_phone* indicates whether the *Core-household* and *Core-individual* interview was conducted over the phone.

The following data and materials are provided. The suffix [date] indicates the version of the data.

Documentation:

- *SMP - User guide - 1 Overview – 20260330.pdf*: User guide introducing the Senegal Migration Panel, its sampling and use across waves. This user guide is not part of the Wave 1 data package, but can be accessed at <https://www.rwi-essen.de/en/research-advice/further/research-data-center-ruhr-fdz/data-sets/microdata/senegal-migration-panel>.
- *SMP - User guide - 2 Wave 1 (2022) – 20260330.pdf*: User guide of the SMP wave 1 (this file).

Village chief survey:

- *smp_2022_chief_20260330.dta*: Dataset
- *smp_2022_chief_questionnaire.xlsx*: Questionnaire
- *smp_2022_chief_codebook_20260330.txt*: Codebook

Census survey:

- *smp_2022_census_20260330.dta*: Dataset (format: wide)
- *smp_2022_census_questionnaire.xlsx*: Questionnaire
- *smp_2022_census_codebook_20260330.txt*: Codebook

Network, Core-household, Core-individual and phone mop-up surveys:

- *smp_2022_core_20260330.dta*: Dataset (format: wide)
- *smp_2022_core_questionnaire.xlsx*: Questionnaire
- *smp_2022_core_phone_questionnaire.xlsx*: Questionnaire (phone)
- *smp_2022_core_codebook_20260330.txt*: Codebook

6.2 Unique identifiers

Households and their members are identified via unique identifier variables that can be used to link observations across surveys and waves:

hhid_chief_anon uniquely identifies the **household** of the village chief in the *Village chief* data.

hhid_anon uniquely identifies the **household** in the *Network, Core-household* and *Core-Individual* data and replaces the variable *hh_id* from the questionnaire.

ind_id_anon_* uniquely identifies each household **member** listed in the *Census*. In the *Network, Core-household* and *Core-individual* data, ***ind_id_anon*** can be used to link respondents of different surveys or survey modules to roster information in the *Census* or across surveys (see Section 6.3 “Identifying and linking individual information and respondents”). To do so, it is best to reshape the *Census* dataset from *wide* (its current format) to *long*.

ea_anon uniquely identifies villages in all datasets and replaces the variable *ea_id* from the questionnaire.

6.3 Identifying and linking individual information and respondents

The following provides information on how to use identifier variables to link individual information and respondents of different surveys.

Village chief survey respondent:

Variable **hhid_chief_anon** is the unique identifier of the household of the village chief. For anonymization, this identifier cannot be linked to the *Census*, *Network*, *Core-household* or *Core-individual* surveys. However, variable *hhid_chief_anon* can be used to link observations of the *Village chief* survey across waves (chiefs were re-interviewed in 2024).

Variable **ea_anon**, which uniquely identifies villages, can be used to link village-level information collected in the *Village chief* survey to households in these villages in the *Census*, *Network*, *Core-household* and *Core-individual* surveys.

Census survey respondent:

Respondents of the *Census* survey are identified via the variables **respondent_anon***. The variables contain the *ind_id_anon* of the respondents and can thus be linked to *ind_id_anon* in the *Network*, *Core-household* or *Core-individual* data (*respondent_anon1* contains the *ind_id_anon* for the first individual listed as a respondent for the *Census*, *respondent_anon2* lists the *ind_id_anon* for the second respondent and so forth). In cases where someone not listed as a household member responded to the *Census* survey, the *respondent_anon** variables are filled with one of the following: neighbor/concession member, child in HH, other family member (not part of HH), village chief, and other individual (relation unknown). Responses are grouped based on responses to the variable *respondent_nonmember* that asks who the respondent was. When only a name without additional information was listed, *respondent_anon** was coded as “other individual (relation unknown)”. *respondent_nonmember* was dropped after assigning responses into the above categories.

Respondents of the *Network* and *Core-household* surveys:

Respondents for the *Network* and *Core-household* surveys are identified via the variables **hh_resp_id_anon*** which can be linked to *ind_id_anon* in the *Census* and *Core-individual* data (*hh_resp_id_anon1* contains the *ind_id_anon* for the first individual listed as a respondent for the household-level component, *hh_resp_id_anon2* lists the *ind_id_anon* for the second respondent and so forth). The maximum number of individuals listed as respondents is 8.

These variables replace the variable *hh_respondent* in the survey instrument that presents a list of all members that were listed in the household roster and asks which household member(s) from the list participated in the interview. For respondents not included in the list of household members, enumerators were asked to record the name of this individual in *hh_respondent_other*. In these cases, *hh_resp_id_anon** contains “other HH member” and *hh_respondent_other* was deleted.

Network-module respondents:

Respondents of the network module in the *Network* survey are identified via the variables **network_resp_id_anon***. The variables contain the *ind_id_anon* of the respondents and can thus be linked to *ind_id_anon* in the *Census* and *Core-individual* data. *network_resp_id_anon** replace the original survey instrument variables *networks_respondents* and *networks_respondents_other*.

Core-individual respondents:

The respondent of the *Core-individual* survey is identified via the variable *ind_id_anon*.

6.4 Variable labels

Variable labels include the following prefixes to improve data usability:

- *VGEN*: Variables that are generated or adjusted and are not part of the questionnaire.
- *RGEN*: Variables that are generated because the origin variable allowed for multiple responses. *RGEN*-variables are dummies indicating if a response category was selected.
- *Calc*: Variables of type “calculate”, calculated within the survey form in SurveyCTO.
- (*Phone*): Variables that are only asked in the phone mop-up of the *Core-individual* and *Core-household* surveys, either because the wording differed from comparable items in the in-person survey, because they were repeated (e.g. enumerator codes) or because these were exclusively phone survey items. These variables also have the variable name suffix *_phone*.
- Variables without a prefix are as they appear in the questionnaire.

7. Recoding of data

7.1 Updating the member roster in the *Census* survey using subsequent data

During the selection of the respondent for the *Core-individual* survey in phase 2 of the data collection (see Section 5.3 “Respondent selection of the *Core-individual* survey”), 373 household members were identified who had not been listed in the previous *Census* with characteristics that made them eligible as *Core-individual* respondent (because of incorrect age or gender, N=191) or had not been listed at all (N=182). In addition, 401 individuals had been identified in the previous *Census* as eligible as *Core-individual* respondent but were not, because they were actually female (N=212) or outside the age range of 18 to 40 (N=189).

The following variables in the *Core-individual* questionnaire provide information about the additional household members, but were dropped from the dataset for usability:

- *respondent2_more*: Is there another household male member aged 18 to 40 who you could interview?
- *respondent2_more_age*: How old is this person?
- *respondent2_more_why*: Why did we not have the record of the person from the roster survey?
- *respondent2_more_member*: Which of these members does the person you selected refer to?
- *respondent2_*_other*: Specify other

Erroneously entered age and gender information in the *Census* was corrected using the information collected during the respondent selection of the *Core-individual* survey.

Household members not listed during *Census* data collection at all were added to the *Census* data. This was the case for 181 household members. Added household members are flagged by the variables *new_member_**. Note that the only available information for these newly added members are age and gender.

Variables *total_adults*, *adult_current*, *adult_roster_count*, and *child_current* are adjusted for these changes. When an individual is reported to be below 15 years and was previously marked as 18-40, the roster entry for this individual still exists. Thus, to obtain household statistics such as total men or women, these individuals should be excluded using an age filter.

7.2 Household head recodes in the *Census* survey

In the *Census* survey, household heads are identified via the variable *relation_roster_**, which asks the relationship of each individual listed in the roster to the household head, and *relation_roster_** == "self" indicates that the member is the household head. There are certain cases in the raw data where either no individual was or multiple individuals were listed as *relation_roster_** == self. In these cases, enumerators and their supervisors were consulted to identify the household head. When available, the correct relationship to the actual head of members who were listed as the head but were not actually the head was entered. A new variable *is_head_** was created, which is a dummy indicator equal to 1 if the household member is the head.

7.3 Migration roster recodes in the *Census* survey

In the *Census* survey, the enumeration of past year internal and external migration behavior of household members (migration rosters) was updated during the survey to simplify and shorten the interview. Initially, the survey instrument asked for each migration episode of at least 4 weeks (this could include multiple episodes to the same destination). This was updated to ask for each migration destination where the member spent at least 4 weeks (the latter question format is reflected in the provided questionnaire). 2,035 households (20.5% of households) reported their internal migration behavior, and 461 households (4.64% of households) reported their external migration behavior, using the old format of migration rosters.⁵ For all households responding to the old format of migration rosters, their migration information was mapped to the updated format of migration rosters. Variables *old_intern_roster* and *old_extern_roster* indicate if the household responded to the old format of migration rosters and therefore whether their information was mapped to the variables of the new format of migration rosters. After the mapping the old roster format variables were dropped, such that the data only includes the new roster format variables.

7.4 Occupation recodes in the *Census* and *Core-individual* surveys

Variables *occupation_** in the *Census* survey and variables *activity_now*, *activity_6m* and *i_occupation_future_entry* in the *Core-individual* survey used the same response list with occupation categories during enumeration. A new response list was coded to include occupations that were recorded numerous times as a response to the "other" variable. The resulting new list with occupations is used in future surveys. Table 4 provides a comparison of the original codes and occupations as they appeared in the questionnaires and the updated codes and occupations as they appear in the data.

Responses to *occupation_**, *activity_now_other* or *activity_6m_other* that indicated the respondent is an apprentice in a given occupation are replaced with the occupation code for students (19).

For anonymization, information on the occupation "village chief" has been removed. All binary variables named *occupation_26_**, indicating the village chief occupation, were deleted. In the multiple response variables *occupation_** the occupation code 26, corresponding to the occupation of village chief, was replaced with -66 (other). If -66 (other) was already present in the multiple response entry, the occupation code 26 was simply removed.

⁵ The numbers differ for internal and external migration behavior as fewer households had any external than internal migration behavior to report.

Table 4: Updated list with response codes for occupations in the *Census* and *Core-individual* surveys

	Original codes (as in questionnaire)	Updated codes (as in data)
-99	Refuse to answer	Refuse to answer
-66	Don't know	Other
0	No occupation	No occupation
1	Farm worker (own field)	Farm worker (own field)
2	Sourga	Sourga (Farm someone else's field)
3	Nurse / midwife	Nurse / midwife
4	Public administrative officer	Public administrative officer
5	Independent retailer	Independent retailer
6	Salaried salesperson	Salaried salesperson
7	Hairdresser	Hairdresser
8	Tailor	Tailor
9	Blacksmith	Blacksmith
10	Mason	Mason
11	Mechanic	Mechanic
12	Carpenter	Carpenter
13	Miller / mill manager	Miller / mill manager
14	Welder / metalworker	Welder / metalworker
15	Driver, motorcyclist, taximan, jakarta	Transporter (driver, motorcyclist, taximan, jakarta)
16	Domestic / maid	Domestic / maid
17	Marabout / healer	Marabout / healer / religious leader
18	Peon (débrouillard)	Peon
19	Student	Student
20	Fisherman	Fisherman
21	Livestock breeder	Livestock breeder
22	Electrician	Electrician
23	Retired	Retired
24	Salesperson	Salesperson
25	Commercant	Commercant
26	Village chief (dropped from data)	Village chief (dropped from data)
27	Guard	Guard
28	Baker	Baker
29		Miner
30		Real estate
31		Farmer (someone else's field)
32		Shoe maker/repairer
33		Teacher
34		Painter
35		Plumber

Source: Own data.

8. Animal recodes in the *Core-household* survey

In the *Core-household* survey, the list with animals that were enumerated was aggregated and shortened. The old format asked how many of each of the following a household owns:

- Animals_horse: Horses
- Animals_donk: Donkeys
- Animals_ox: Oxen
- Animals_goat: Goats
- Animals_rabbit: Rabbits
- Animals_sheep: Sheep
- Animals_chick: Chicken
- Animals_other: Other

These items were condensed into the following items:

- Animals_draft: Horses/donkeys/oxen (draft animals)
- Animals_cattle: Cows/bulls (cattle)
- Animals_ruminants: Sheep/goats (small ruminants)
- Animals_poultry: Chicken (or other poultry)
- Animals_pigs: Pigs

Responses from the old format were recoded to match the new format. For example, the sum of *animals_horse*, *animals_donk*, and *animals_ox* is replaced as the value for *animals_draft*. The original variables were deleted from the final version of the data.