

World Bank Household Surveys for the Africa Migration Project in Burkina Faso

Report: Methodology, survey implementation and data entry

The main objective of the assignment is to contribute to the improvement of migration and remittances impact on development in sub-Saharan Africa. Some specific objectives are to provide better understandings of migration and remittances through household surveys. By enhancing the knowledge regarding migrant remittances sending or receiving, these data will be a reliable input for policy-makers. The survey will provide information on migrant and non-migrant households' socioeconomic characteristics and how they send remittances back home. Appropriate approach and methodology will be applied to carry out the survey activities for high quality data.

1. Sampling Approach

The sampling approach was chosen to focus on international migration and remittances. It has ensured an adequate representation of migration density in Burkina Faso. First we used the Burkina Faso Census 2006 that includes questions on demographic and migration. Then we calculated the incidence of international migration for each of the 45 provinces in Burkina Faso. We chose the largest 10 provinces which have the highest international migration incidence. The universe for this survey is the households that are placed in these provinces. After chosen the universe of households, we have adopted a two stages sampling: i) village/town sector and ii) household. Migration scope is different depending on the province or rural/urban area in Burkina Faso. We determine the number of village to be selected in each province according to the international migration incidence of each province. In total we selected 78 villages out of 2,273 villages for the 10 provinces.

Table 2: Migration incidence, number of villages and population for selected provinces

Province	International migration density/incidence	Number of villages/sectors	Population
Sanmatenga	11.7	386	598,014
Yatenga	4.4	424	553,164
Passoré	2.9	220	323,222

Boulgou	2.9	314	543,570
Boulkiemdé	2.8	175	505,206
Namentenga	2.6	176	328,820
Banwa	2.6	118	269,375
Sourou	2.5	159	220,622
Tuy	2.5	102	228,458
Kadiogo	2.4	199	1,727,390
Sub-Total for 10 provinces	3.0	2273	5,297,841
Burkina Faso	1.2	8245	14,017,262

Source: Institut National de la Démographie et la statistique (INSD), Burkina Faso, 2006.

Village town sector selection

The first stage of sampling consisted of villages and towns selection. We randomly selected the village/town sector within each province with equal probability. We have considered a village/town sector sample size of **27 households**. We distributed the 27 households according to the following rule: 9 households with international migrants, 9 households with internal migrants and 9 households with no migrants. Thus, the whole sample is composed of **78 villages/towns sectors** for **2,106 households**. Considering migration density in each province, a number of villages/town sectors were randomly selected by province. In big town as Ouagadougou, each sector was considered as a village.

Table 2: Number of villages and households selected by Province

Province	Number of villages / town sectors	Number of Households listed	Number of Households interviewed
Sanmatenga	24	2734	648
Yatenga	9	1139	243
Passoré	6	590	162
Boulgou	8	876	216
Boulkiemdé	6	926	162
Namentenga	6	936	162
Banwa	5	639	135
Sourou	5	599	134
Tuy	5	323	135
Kadiogo	4	435	108
Total	78	9177	2102

Source: World Bank Migration Survey-Burkina, 2009.

Household selection

The second sampling stage was household selection. Households were randomly selected after doing an exhaustive and quick household census in the selected village. For the villages in rural areas, we listed all the households in the village. The listing survey was not done for all households in town: we have randomly selected about 120 households in each selected town sector around a random start point. This number is extended if needed in order to attend the household type quotas. We have determined this point by using the town addressing codes. The process consisted of selecting randomly a street code and an avenue code which constituted the two elements of the start point. The main objective of the census is the household sampling. It provides information on the head household name, household size and household category (households without migrants, households with internal migrants and households with international migrants). The sub sample size for each household category is 9 in selected village/town sector. These households are randomly selected among the total households of the category from the quick census.

2. Survey Implementation

This section describes the *i*) field work team organisation, *ii*) survey supervision and *iii*) data entry and cleaning.

Field work team organisation: The household survey covered 78 villages/town sectors in 10 provinces. That has needed 30 interviewers and 10 supervisors for 10 field teams. Their selection was based on best-experience in conducting household survey. They come from different regions of Burkina with different local languages (at least 5). Most of the supervisors have at least 5 years experience in household survey. Interviewers and supervisors were trained according to the questionnaire and other delicate social considerations to be highlight in such kind of survey. Survey manuals were prepared for interviewers and supervisors. Then, each team was composed of one supervisor and 3 interviewers, taking account of the local language requirements. All the team members (interviewers and supervisor) worked and moved together from village to village. This team organisation has allowed efficient peer monitoring and assistance, and created some dynamism during the survey. The field work has been done from October 29 to November 25, 2009.

Survey supervision: The supervision is at the core of survey data quality. It was assured firstly by supervisor at the team level. The supervisor is the field team leader and his main task is to verify and insure that data are well collected by interviewers. He is the main technical assistance provider in the survey field and then, he is responsible for household sampling in the village. This random household sampling was conducted at the presence of population in the village or village representative persons. The supervisor has presented his team to local representative

persons. He organised his team to be efficient in survey field. He has also managed any problem within team or between team and households. He has informed the survey staff about problems he could not manage. This report has informed us about problems his team has encountered, how it has been solved, and what consequences for data quality and analysis.

A global supervision for all the teams was done by the survey staff. It consisted of improvised team visiting. Each team was visited 2 times. Other supervision visits was needed for urgent situations as those encountered in the Boulgou Province. The main objective was to insure the correct survey ways for all teams in the work field and to share efficient astuteness found by any team for best quality data. We have focused a particular attention on the first global supervision in order to correct some possible survey errors at the earliest.

We have encountered few difficulties in implementing this survey. Problems have been solved with less significant effect on data quality. For example, it was difficult to estimate the value of some goods that are not common or not found at the local market. Some interviewed household members were not willing to give information on non household members from which they have received transfers. To convince particularly these members for anonymous, it were not asking them the name of sender; the interviewer just wrote a surname. The reticence of interviewed household members to participate to the survey was higher in urban areas. The insistence in explanation of the survey purpose and the time efficiency of the interview have been the main strategies we have adopted to solve this problem. The collaboration with a few local authorities was not good as expected, probably because of political reason. However, the team provided the information to local authorities to conduct the survey.

Most of the questionnaires (99%) have been completed at the first visit. The non response rate is very low: 4 households out of 2106 have not responded. This score could be explained by the census process during which the high experienced team prepared the household for interview if it coma to be selected.

3. Data entry, cleaning and basic statistic analysis

Data have been entered by 12 persons, using EpiData software. These persons have a high experience and most of them have understood the questionnaire. They have been trained for this data entry. We have set up an appropriate data entry grid with rigorous consistency checks. The data entry was supervised by the survey staff. Data have been cleaned using STATA software. The double entry process was the first level of dataset quality control. Several consistency and cross checking rounds have been done in order to ensure a high dataset quality. This process often needed to check error using the hard completed questionnaires. Very few errors have been unsolved. We made no imputation for missing data in the dataset.

4. Weight calculation

Following the sampling frame, weights are calculated at three levels: province, village and household. The province is weighted considering its population and the total population of the survey area (10 provinces). The village is weighted in the same way, according to its population, the total population of the corresponding province and the total population of the 10 provinces of the survey area. As households have been selected in three categories (households with no migrant, households with internal migrant and households with international migrant), the weight of each household is calculated taking into account the number of households in its category. It also includes the weights of its village and province. For any level, the weight is the inverse of the level unit selection probability. Further calculation details are in the STATA do-file programs while weights are in the database from the survey.

We have adopted a probabilistic two stage sampling method. It is necessary to correct the inequality in probability for the observed unit (household) to be selected in the survey universe. The standard correction method consists of weighting the statistic outcomes. The sampling weight (w_{si}) for a sample unit i in a strata s is calculating by inverting its probability to be selected

$$w_{si} = \frac{1}{\varphi_{si}}$$

where φ_{si} is the probability for the unit i in the strata s to be selected.

In this survey, the probability φ_{hj} for the household h from the type of household j in the village v of the province k to be selected in the survey universe is:

$$\varphi_{hj} = \frac{pop_k}{pop} \cdot \frac{nvs_k}{nv_k} \cdot \frac{nhs_j}{nh_j}$$

where:

pop_k : population of the province k ;

pop: total population of the 10 provinces representing the universe;

nvs_k : number of villages selected from the province k ;

nv_k : total number of villages of the province k ;

nhs_j : number of households selected from the strata j of household;

nh_j : total number of households of the strata j of household;