

LIBERIA
NATIONAL HOUSEHOLD FOREST SURVEY
2018-2019

BASIC INFORMATION DOCUMENT

Prepared by the World Bank
Version: August 2020

ABBREVIATIONS

CAPI	Computer Assisted Personal Interviewing
CIFOR	Centre for International Forestry Research
EA	Enumeration Area
FAO	Food and Agriculture Organization of the United Nations
FDA	Forestry Development Authority
HIES	Household Income and Expenditure Survey
HH	Household
LSMS	Living Standards Measurement Study
LSMS-ISA	Living Standards Measurement Study – Integrated Surveys on Agriculture
LISGIS	Liberia Institute of Statistics and Geo-Information Services
NHFS	National Household Forest Survey
PES	Payment for ecosystem services
PHC	Population and Housing Census
PROFOR	World Bank Program on Forests
RuLIS	Rural Livelihoods Information System

TABLE OF CONTENTS

ABBREVIATIONS	2
1. INTRODUCTION	5
2. SURVEY DESIGN	5
2.1. QUESTIONNAIRE DESIGN	5
2.1.1. HOUSEHOLD QUESTIONNAIRE	6
2.1.2. COMMUNITY QUESTIONNAIRE	8
2.2. SAMPLING DESIGN	9
2.3. DEFINITION OF FOREST	9
2.4. FIRST STAGE SELECTION	9
2.5. SECOND STAGE SELECTION	10
2.6. POST-FIELDWORK ADJUSTMENT	11
3. ORGANIZATION OF THE SURVEY	14
3.1. SURVEY MANAGEMENT TEAM	14
3.2. TRAINING OF FIELD STAFF	14
3.3. COMPOSITION OF FIELD TEAMS	14
3.3.1. SUPERVISORS	14
3.3.2. ENUMERATORS.....	15
3.3.3. MAPPERS & LISTERS.....	15
3.4. FIELDWORK MONITORING AND EVALUATION	15
4. DATA ENTRY AND DATA MANAGEMENT	15
4.1. DATA ENTRY PLATFORM & DATA MANAGEMENT	15
4.2. DATA CLEANING	15
5. USING THE DATA	16
5.1. FILE STRUCTURE.....	16
5.2. KEY IDENTIFIERS	17
5.3. CONFIDENTIAL INFORMATION	18
6. WEIGHTING.....	18
7. INCOME AGGREGATE	20
REFERENCES	24
ANNEX I. SAMPLE SIZE CALCULATIONS	25
ANNEX II. DATA FILE STRUCTURE	27

LIST OF TABLES

Table 1. Contents of the Household Questionnaire	6
Table 2. Contents of the Community Questionnaire.....	8
Table 3. First Stage Sample	10
Table 4. NHFS Sample by County.....	10
Table 5. Cluster Composition	12
Table 6. NHFS Sample by Cluster.....	13
Table 7. Sample Size Estimates	26
Table 8. Structure of HH Data Files	27
Table 9. Structure of Community Data Files	28
Table 10. Structure of Supplemental Data Files	28

LIST OF FIGURES

Figure 1. Distance from Forest to Selected EAs	12
Figure 2. Forest Cover of NHFS Sample	13

1. INTRODUCTION

The World Bank and the Liberian Government, through the Liberia Institute of Statistics and Geo-Information Services (LISGIS) and the Forestry Development Authority (FDA), have undertaken the first National Household Forest Survey in Liberia. The National Household Forest Survey (NHFS) is the first of its kind in Liberia and is intended to address a large data gap related to quantification of the contribution of forests and forest products to the livelihoods of rural communities, in a systematic and statistically robust manner. The survey aims to collect socioeconomic data on households that live in or near forested areas, as well as data on the forest products upon which they rely for their welfare and livelihoods. The purpose of this document is to provide an overview of the implementation of the 2018 National Household Forest Survey in Liberia and the resulting data.

The survey was funded by the World Bank through its Forest Carbon Partnership Facility REDD+ Readiness Support grant, implemented by the FDA. Fieldwork was implemented by LISGIS, with technical support provided by the World Bank's Living Standards Measurement Study (LSMS) team, generously funded by the Program on Forests (PROFOR), and the Environment and Natural Resource Global Practice.

After intensively training field staff and pilot-testing the questionnaire, the survey was fielded from September 2018 to January 2019. Data collection was undertaken at a national scale for areas falling within 2.5km of forests, from 250 Enumeration Areas (EAs) spread across Liberia's 15 counties (excluding urban Montserrado). The survey consisted of two parts: a household questionnaire that covered 3000 households, and a community-level questionnaire that was administered in each of the 250 selected EAs.

2. SURVEY DESIGN

2.1. QUESTIONNAIRE DESIGN

The NHFS survey instrument was based on the publicly available National Socioeconomic Surveys in Forestry guidebook and set of specialized forestry modules (FAO et al., 2016).¹ These tools were developed jointly by the Living Standards Measurement Study (LSMS) team of the World Bank, the World Bank's Program on Forests (PROFOR), the Food and Agriculture Organization of the United Nations (FAO), the Center for International Forestry Research (CIFOR), and the International Forestry Resources and Institutions Network.

The guidebook modules, originally designed for universal adoption, were adapted to the Liberian context. Borrowing from the Liberia Household Income and Expenditure Survey (HIES) instruments², they were then supplemented with several modules on income to allow for

¹ National socioeconomic surveys in forestry: Guidance and survey modules for measuring the multiple roles of forests in HH welfare and livelihoods (FAO. 2016). Downloadable at: <http://www.fao.org/forestry/forestry-modules/en/>.

² LISGIS 2016.

computation of total HH income. In addition, the NHFS team developed a questionnaire module on gender-related aspects of forest enterprises and forest-related community participation.

The NHFS survey consisted of:

1. A HH questionnaire, administered to 12 selected HHs in each enumeration area, and
2. A community questionnaire, administered to a group of members from the EA.

Each questionnaire was administered using computer-assisted personal interviewing (CAPI) with CSPro³ software. While the forestry modules are available in the World Bank's Survey Solutions CAPI software, LISGIS is familiar with CSPro and preferred implementing the NHFS with their current program.

2.1.1. HOUSEHOLD QUESTIONNAIRE

The household questionnaire is a multi-topic survey instrument that includes twenty-one sections covering a wide range of topics, such as economic activities, demographics, welfare and other forest-related information (see Table 4 below).

Table 1. Contents of the Household Questionnaire

Module	Description
Section 1: Household identification & Survey staff details	This module lists the household identifiers, information on household location, date of interview, supervisor and enumerator codes.
Section 2: Household member roster	This module contains the roster of individuals living in the household, their gender, age, relationship to the household head, duration away from the household in past 12 months, number of days meals were taken in the household, where born, how long in this community, and information on the location and level of education of parents of every member, including ID's if in the household.
Section 3: Forest resource base	This module contains basic information on the distance of the household from the forest, the change in forest extension and the reasons of this change.
Section 4: Forest benefits	This module contains information on the forest services and benefits observed by the household.
Section 5: Forests and health	This module contains information on the use of forestry products for health care and medical assistance.
Section 6: Forests and energy	This module contains information on the use of forestry products (fuelwood and charcoal) for cooking, heating, and lighting.
Section 7: Income from forests (collected products)	This module contains information on income generated from and consumption of collected forest and wild products.
Section 8: Income from forests (processed products)	This module gathers information on income generated from and consumption of processed forest and wild products.
Section 9: Labor income	The module is administered to all individuals 10 years or older and collects labor information over the last 12 months. This module contains on participation in work activities as an

³ See: <https://www.census.gov/data/software/cspro.html>

	employee for a wage, salary, commission or any payment in kind (including doing paid apprenticeship, domestic work or paid agriculture or forestry); run/operate a non-farm business of any size, help in any kind of non-farm business run; household farm activities (including raising livestock or producing crops whether for sale or for household food); type of primary and secondary work, employers and wages; participation in unpaid apprenticeships, and other unpaid labor.
Section 10: Land parcels	This module contains information on cultivated or owned household land, such as ownership, use, types of crops and cultivation, and types of fertilizers.
Section 11: Crop disposition	This module contains information on all crops grown by the household in the last 12 months, including prices per unit of total production, consumption, in kind payments, and gifts.
Section 12: Household non-farm enterprises	This module contains information on non-agricultural family enterprises or trading business, specifically who manages/owns the enterprise, employees, enterprise operation periods, start-up capital and source, customers, business trends, sales revenue, expenditures, and profits. Non-farm enterprises exclude farm enterprises that relate to crop production, while they include forest-related business and other agricultural-related enterprises, such as selling produce at market, hunting and selling meat from hunted animals, and fishing. Forest-related enterprise activity should only be reported in Section 12 (it should not be also in Section 7 or 8).
Section 13: Other forest income including PES	This module contains information on other forest-related income sources, including payments for environmental services programs.
Section 14: Other non-labor income	This module contains information on other non-labor income, such as different forms of incoming transfers or gifts, pension & investment income, rental income, and revenue from sale of assets.
Section 15: Household assets	This module contains information on household assets, including the number, the price, and use for forest-related activities.
Section 16: Food security	This module contains information on household food security over the last 12 months and the importance of the use of forestry products during times of insecurity.
Section 17: Shocks and crises	This module contains information on the negative events that affected the household over the last 12 months and the importance of the use of forestry products during the critical months.
Section 18: Household access	This module contains information on the distance of the household from the nearest health center, school, main road and market.
Section 19: Forests and construction	This module contains information on the use of forest products for construction.
Section 20: Forest clearance	This module contains information on the forest area cleared, abandoned and planted over the last 5 years and last 12 months.
Section 21: Household recontact information	This module contains phone numbers of the head of the household and partner, time and day of the end of the interview.

2.1.2. COMMUNITY QUESTIONNAIRE

The community questionnaire includes nine sections that gather information on a range of community characteristics, including community level forest access, forest-related programs, and prices (see Table 2 below). A “community” is defined as the village of the enumeration area selected for inclusion in the sample and in which most residents recognize as being their community. The community questionnaire was administered to each community associated with the 250 EAs, through a facilitated focus group approach. Community members present in the community interviews included a diverse selection of knowledgeable residents such as the village headman, the headmaster of the local school, etc. Every community interview group included both men and women.

Table 2. Contents of the Community Questionnaire

Module	Description
Section 1: Cover	This module lists the community identifiers, information on community location, date of interview, supervisor and enumerator codes.
Section 2: Roster of Informants	This module lists the group of informants and their age, sex, positions in community, length of residence in the community, education and language spoken.
Section 3: Forest Roster	This module collects basic characteristics of the forests by community, such as the name, management type, year rules and regulations were established and forest size.
Section 4: Seasonal Calendar	This module contains information on months of the forest and wild product collection and sale.
Section 5: Community Forest Status	This module contains information on the total area of accessible forests, forest clearance, and forest protection agreements.
Section 6: Most Important Forest and Wild Products	This module contains information on the 3 most important forest and wild products for income generation and subsistence separately.
Section 7: Units and Pricing	This module contains information on standard and non-standard units and prices for forest and wild products in the community.
Section 8: Community Participation and Support for Forest-Related Activities	This module contains information on the benefits received by the community from the participation in forest-related programs over the last 5 years (i.e. sustainable use of forest, conservation of parts of forests for biodiversity, conservation of parts of forests for watershed protection, forest fires and pest control practices, grazing management, and permitting access to forest); and the benefits received by the community from the external forest-related support over the last 5 years (i.e. technical assistance for forestry practice; training in forest management; information about forest policies and laws; training in forest product processing; free seedlings; free implements for forestry operations; and free growth/protection inputs for forests)
Section 9a: Gender Activities (Male respondents)	This module collects information <u>only from male respondents</u> on not only the forest-related activities that women participate in and the roles they play, but also on their involvement in the decision-

	making processes regarding forest-related policy and structure within the community.
Section 9b: Gender Activities (Female respondents)	This module collects information <u>only from female respondents</u> on not only the forest-related activities that women participate in and the roles they play, but also on their involvement in the decision-making processes regarding forest-related policy and structure within the community.

2.2. SAMPLING DESIGN

The original sampling strategy, which employed a stratification on the basis of distance from EA to nearest forest, was revised following fieldwork upon realization of an error in the computation of the distance to forest which was used to stratify EAs. The sections below discuss the original strategy as well as the measures taken to rectify the erroneous stratification.

DEFINITION OF FOREST

Given the focus of the NHFS on the population living in close proximity to forests⁴, a first step was to clearly define *forest* for the purposes of the survey. Building on the national definition of forest used in Liberia, and modifying it in order to minimize the impact of small urban forests and facilitate survey operations, the NHFS employed the following definition:

Forest = area with at least 30 percent tree canopy cover, with trees higher than 5 meters and at least 50 hectares in size⁵

The forest cover was determined using high-resolution forest cover data produced in 2019 based on satellite information on forest cover in Liberia for 2015.⁶ All EAs within 2.5 kilometers of forests identified with this definition were deemed eligible for inclusion in the NHFS.⁷ EAs from the Montserrado county (part of Greater Monrovia) were excluded from the sample universe due to the high rate of urbanization. However, rural parts of Montserrado county were included in the sample universe.

FIRST STAGE SELECTION

Liberia is divided administratively into 15 counties. Each county is divided into districts, which are further subdivided into clans. For the purposes of statistical surveys, clans are further divided into small operational areas, known as Enumeration Areas (EAs). There are about 7,000 EAs in the country.

The initial sampling strategy, which was used to select the primary sampling units (EAs) for the NHFS, was designed to stratify EA selection by proximity to forest. Based on the forest definition

⁴ The survey focused on forest-proximate HHs as these would likely be the most dependent on forest and also the HHs for which the contribution of forests to livelihood and incomes would likely be most imperfectly measured in traditional data collection approaches.

⁵ In addition, any forest patch with a perimeter to area ratio more than 0.02 was excluded. This restriction was imposed to focus on non-fragmented and relatively large forest areas capable of providing both consumptive and non-consumptive goods and services.

⁶ Metria and Geoville. 2019

⁷ Distance from enumeration area to the nearest forest is computed from the centroid of the enumeration area.

defined above, the distance from each EA in the country (except urban Montserrado) to the nearest forest was computed. That distance was subsequently used to assign each EA to one of the following strata: S1 (less than 2km from forest); S2 (two to 7 km from forest); S3 (7 to 15 km from forest).

Following strata classification, a total of 250 EAs were selected through a Probability Proportional to Size (PPS) sampling approach within each stratum, with the following purposeful allocation across strata: 90 EAs in S1; 90 EAs in S2; 70 EAs in S3.⁸ The measure of size for each EA was based on the total number of households listed in the 2008 PHC.

Table 3. First Stage Sample

Original Strata	No. of EAs in Strata	No. of EAs Selected
S1 (<2 Km from forest)	733	90
S2 (>2 but <7 Kms)	375	90
S3 (>7 but <15 Kms)	236	70
Total	1,344	250

SECOND STAGE SELECTION

Following the selection of the 250 sample EAs, a listing of households was conducted in each sample EA to provide the sampling frame for the second stage selection of households. Random sampling was used to select 12 households from the household listing for each sample EA. The total sample target of 3,000 households was derived based on the sample size calculations presented in Annex I, and with consideration for budgetary constraints. While the original sample design provided a total household sample size of 3,000 (250 EAs with 12 households sampled per EA), data from 14 households are missing or unusable, representing 0.05 percent of the sample and resulting in a final sample of 2,986 households. Similarly, data from 5 of the community questionnaires were missing or unusable, resulting in a total sample of 245 community questionnaires. The final sample of 2,986 households is distributed across countries as illustrated in Table 4.

Table 4. NHFS Sample by County

County	HH Count
Bomi	168
Bong	288
Gbarpolu	300

⁸ This is a well-accepted approach to sampling, widely used in surveys, including for other LSMS surveys. A succinct explanation of the approach, and its application (including calculation of the basic probability weights) can be found at:

http://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/prevalence_survey/psws_probability_prop_size_bierrenbach.pdf

Grand Bassa	370
Grand Cape Mount	204
Grand Gedeh	216
Grand Kru	108
Lofa	264
Margibi	120
Maryland	108
Montserrado	36
Nimba	275
River Gee	108
Rivercess	275
Sinoe	146
Total	2,986

POST-FIELDWORK ADJUSTMENT

Upon post-data collection analysis, it was discovered that the initial variable that was used to stratify EAs by distance to forest was incorrectly computed. Despite thorough attempts to understand the nature and source of the error, it was determined that a mechanical error must have occurred during the process of the distance calculations. This error rendered the stratification incorrect. Therefore, the stratification by distance to forest has been abandoned and the sample weighted to reflect only geographic clusters, not distance to forest. This was determined to be the most appropriate way forward following consultation with sampling experts.

The resulting sample, therefore, is weighted to reflect all EAs in Liberia (with the exception of urban Montserrado) that fall within 2.5 km of the nearest forest, which was the upper bound of the distances for the selected EAs. The distribution of the distance from the center points of the selected EAs to the nearest forest, as recomputed by the World Bank following data collection, is presented in Figure 1.

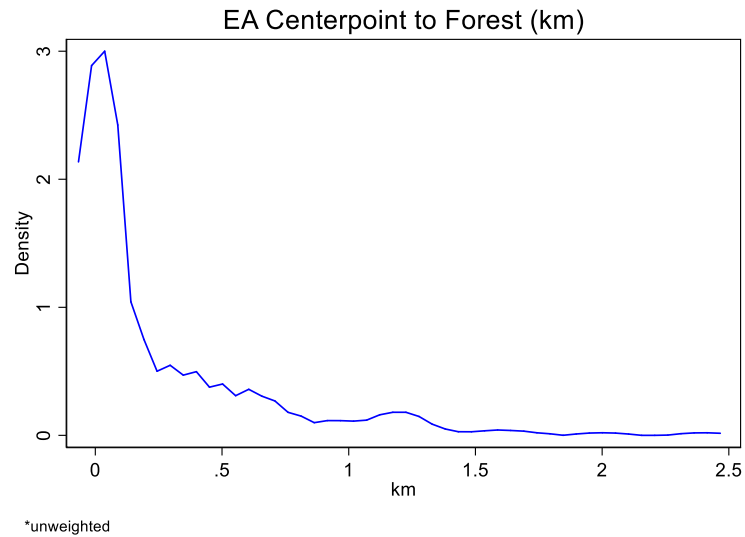


Figure 1. Distance from Forest to Selected EAs

In lieu of the distance stratification, the data has been weighted to reflect geographic clusters (weighting is discussed in detail in Section 6). These clusters were defined based on similarity in the intrinsic and exogenous factors relevant to the interaction between HHs and forests, such as the level of overall socioeconomic development, the extent of forest cover, and the vulnerability to forest loss and degradation. Each cluster includes multiple counties, as illustrated in Table 5. The NHFS dataset includes only the cluster variable (*reg_cluster*), not the original stratification variable. The distribution of households across the clusters is reported in Table 6.

Table 5. Cluster Composition

Clusters	Counties
Western Cluster	Bomi Gbarpolu Grand Cape Mount Lofa
Central Cluster	Bong Grand Bassa Margibi Rural parts of Montserrado Nimba
Eastern Cluster	Grand Gedeh Grand Kru Maryland River Cess River Gee Sinoe

Table 6. NHFS Sample by Cluster

Cluster	No. of Sample EAs	No. of Sample HHs
Western	78	936
Central	91	1,089
Eastern	81	961
Total	250	2,986

Finally, the map provided in illustrates the degree of forest cover in each of the three clusters. The NHFS sample is weighted to reflect the population in all areas of the map except for the areas in white, which represent areas more than 2.5km from an EA center.

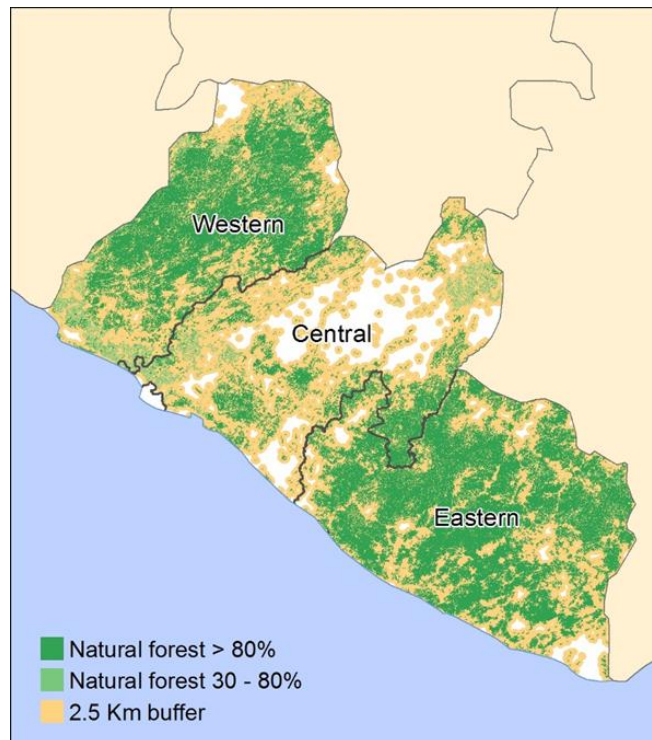


Figure 2. Forest Cover of NHFS Sample (source: author's compilation using Metria and Geoville 2019 data).

3. ORGANIZATION OF THE SURVEY

3.1. SURVEY MANAGEMENT TEAM

The survey was executed by the Liberia Institute of Statistics and Geo Information Services (LISGIS), under the overall direction of its Director General and with technical assistance from the World Bank. The survey management team included staff from the FDA-Liberia, LISGIS, LFSP, and the World Bank's LSMS (Development Data Group) and Environment and Natural Resource Global Practice. The survey management team was responsible for questionnaire design, recruitment of personnel, training of personnel, sampling and implementation of the survey. The World Bank, with FDA, led on the questionnaire design and sampling design, while LISGIS led on the execution of the sample selection and the implementation of fieldwork.

3.2. TRAINING OF FIELD STAFF

The training of field staff, including enumerators, mappers, listers, supervisors and monitors, was held from August 14-23, 2018, at LISGIS HQ. Phase 1 of the training focused on familiarizing the participants with the paper versions of the household and community questionnaires; Phase II focused on familiarizing participants with the CAPI version of the questionnaires. Training also included a field-day for pilot-testing the questionnaires, to gather inputs from respondents and adjust the questionnaire based on feedback from the piloting. The World Bank team worked closely with LISGIS and with two designated staff members from FDA on the training and on the refinement of the questionnaire in CAPI. To be able to accommodate the CAPI changes, the training was paused for two days and it finally concluded on August 25th, 2018.

From the pool of 44 enumerators invited for the training, a final group of the top 35 were selected competitively. This selection was based on their scores on a written test and on an assessment of their performance in the field during the piloting.

3.3. COMPOSITION OF FIELD TEAMS

Seven teams were deployed to the field for data collection. Each team consisted of eight professionals—field supervisor (team-leader), mappers/listers (2) and enumerators (5). These teams were supported in their day-to-day operations by several technical staff at the LISGIS headquarters in Monrovia.

3.3.1. SUPERVISORS

Each team had one supervisor who travelled with the team at all times. The team supervisor is the overall coordinator and supervisor for the fieldwork, as well as the final responsible for ensuring the team is completing all duties in a timely and high-quality manner. If the enumerators have any problems or need assistance, they report to the team supervisor. The team supervisor may also conduct household interviews when necessary in order to keep fieldwork on schedule.

3.3.2. ENUMERATORS

The enumerators are responsible for accurately and completely administering household and community questionnaires by locating assigned households, using GPS technology to mark and record household locations, administering the household and community questionnaires, and finally communicating any problems to the supervisor.

3.3.3. MAPPERS & LISTERS

Mapping and listing personnel were deployed in advance of the enumerators. Together, the mapper and lister mapped the layout of each EA and completed the household listing form. The household listing form was subsequently used for the random selection of households in each EA.

3.4. FIELDWORK MONITORING AND EVALUATION

Field-monitors from LISGIS HQ visited field-teams randomly throughout the duration of the fieldwork to ensure the fieldwork is moving along on schedule and serve as quality control for data collection. Representatives from the World Bank also performed monitoring visits upon launch of fieldwork activities. The LISGIS HQ supervision teams supported field supervisors and enumerators, especially with administrative issues, equipment issues, and questions on the implementation of the questionnaire.

The data was collected via CAPI using CSPro software. That meant that the questionnaires and forms were electronic and the enumerators entered the responses and measurements directly on the computer tablets. The supervisor reviewed the interviews on the tablets prior to sharing the data with HQ.

4. DATA ENTRY AND DATA MANAGEMENT

4.1. DATA ENTRY PLATFORM & DATA MANAGEMENT

To ensure data quality and timely availability of data, the survey was implemented using CAPI. CSPro CAPI software was used given LISGIS' experience with the program. To carry out the survey, a laptop computer and a wireless internet router were assigned to each team supervisor. Each enumerator had a tablet which was set up with the CSPro CAPI application. The use of CSPro allowed for the quicker availability of data relative to paper-based interviewing, as no additional data entry stage was necessary. The enumerators shared their data with the team supervisors who periodically shared the team data with LISGIS HQ. LISGIS HQ then exported the CSPro files to Excel and Stata formats and shared those with the World Bank. Upon receiving updated data files from LISGIS HQ, the World Bank executed a Stata program designed to identify errors in the data and would share with LISGIS an error report.

4.2. DATA CLEANING

The data cleaning process was done in several stages over the course of fieldwork and through preliminary analysis. The first stage of data cleaning was conducted by the field-based teams during the interview itself utilizing error messages generated by the *CSPro* application when a

response did not fit the rules for a particular question. For questions that flagged an error, the enumerators were expected to record a comment within the questionnaire to explain to their supervisor the reason for the error and confirming that they double checked the response with the respondent.

The second stage occurred during the review of the questionnaire by the supervisors. Prior to sharing data with LISGIS HQ, the supervisor was to review the interviewers. Depending on the outcome, the supervisors can either approve or reject the case. If rejected, the case goes back to the respective enumerator and a re-visit to the household may be necessary. Additional errors were compiled into error reports by the World Bank and LISGIS HQ that were regularly sent to the teams and then corrected based on re-visits to the household.

The last stage involved a comprehensive review of the final raw data following the first and second stage cleaning, after data collection completion. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) outliers. However, special care was taken to avoid making strong assumptions when resolving potential errors. Some minor errors remain in the data where the diagnosis and/or solution were unclear to the data cleaning team.

The first and the second stage of the cleaning activities were led by LISGIS and the World Bank provided technical assistance. The third stage of data cleaning was performed by the World Bank team exclusively.

5. USING THE DATA

It is strongly recommended that the end user of the NHFS data familiarize themselves with the questionnaires and manuals. The naming of NHFS data files follows the instrument name and module lettering as listed in the questionnaires and variable names, whenever possible, reflect question numbers as presented in relevant modules. Variable labels, whenever possible, match the question asked in the questionnaires. In some cases, it was necessary to modify the variable labels and cross-referencing the questionnaires will be necessary for accurate use of the data.

To increase the efficiency with which the survey instruments were administered, the NHFS instruments make extensive use of skip patterns. The end users of the NHFS data must be aware of these skip patterns to properly interpret the data. When referencing the available paper questionnaires note that skip patterns are, in most cases, clearly identified by an arrow followed by a number in parentheses (>> 2). The skip codes are explained in detail in the Enumerator Manual.

5.1. FILE STRUCTURE

The file structure of the data directly reflects the modules in the questionnaires. Where modules in the questionnaire contain data with multiple levels of observation, data files have been divided with additional numeric labels. It is recommended that end users of the survey data refer to the questionnaires and manuals when using the data. The index of data files, along with key identifiers

relevant for merging data from different modules, are presented in Table 8 and Table 9 in Annex II.

The data files follow an intuitive naming scheme for easy use by the end user. Each file name gives reference to the instrument component. For example, in the Household folder, file “sect1_public” refers to Household Section 1 that includes household identifiers, information on household location, date of interview, supervisor and enumerator codes. Similarly, in the Community folder, file “comm_sect1_public”, for example, refers to Community Section 1 that includes community identifiers, information on community location, date of interview, supervisor and enumerator codes. Some modules contain sub-sections with filter questions, such as file “sect7_filter_public” in the Household folder, for example, that includes filter questions related to Household Section 7.

Additionally, two supplementary files are provided. The first is a household level income aggregate file (Liberia_NHFS_HH_Income.dta). The methodology for constructing the income aggregate is discussed in Section 7. The second is a file which provides the distance from household to nearest forest, using two different forest definitions (Liberia_NHFS_HH_DistanceToForest.dta). The distance from household to nearest forest is provided for the definition of forest provided in Section 2.3 (i.e., >30% canopy, 50ha min area, and p2a threshold), variable *d_fc1*, as well as for a second definition of forest that uses 80% canopy cover instead of the 30% used in the original definition (variable *d_fc2*).⁹

The file structure for these supplemental data files are found in Table 10 in Annex II.

5.2. KEY IDENTIFIERS

The cover sheet for the household questionnaire (“sect1_public”) captures information on the location of the observation, district, clan and the specific EA, as well as other observation level identification, for example, household identification for the household instrument. The variable “**hhid**” has been created to uniquely identify each household. The variable “**ea_unique**” is unique for each EA and is made up of the county, district, clan and EA codes. For example, the unique enumeration identification, “602002012” is the combination of county variable (“60”), district variable (“20”), clan variable (“02”) and EA code (“012”).

⁹ Metria and Geoville (2019) was used for forest cover. The distance from household to nearest forest was computed using the GPS location of the household captured during the interview. In cases where the GPS coordinates were not captured during the interview, the coordinates captured during the household listing operation were used. If both the interview and listing coordinates were unavailable or deemed to be unreliable, EA center points or averages were used.

5.3. CONFIDENTIAL INFORMATION

To maintain the confidentiality of NHFS respondents, certain parts of the database have not been made publicly available. The confidential variables pertain to (i) names of the respondents to the household and community questionnaires, (ii) village and constituency names, (iii) descriptions of household dwelling, (iv) phone numbers of household members and their reference contacts, (v) GPS-based household locations, (vi) names of the children of the head/spouse living elsewhere, (vii) names of the deceased household members, (viii) names of individuals listed in the network roster, and (ix) names of field staff.

6. WEIGHTING

In order to analyze the data and produce estimates for the clusters discussed in Section 2.6, the sample variables must be weighted using the household sampling weights provided in “sect1_public.dta” as **hhweight**.

The weights have been calibrated to reflect the population residing in all EAs within 2.5km of the nearest forest, for the clusters defined in Section 2.6. The basic weight for each sample household is equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage), given the original sampling strategy. The errors with the initial stratification are addressed in the calibration step. The NHFS sample EAs were selected within each strata with PPS from the 2008 PHC frame. At the second stage, 12 sample households were selected with equal probability from the listing for each sample EA. Therefore, the overall probability of selection for the NHFS sample households can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}}, \text{ where:}$$

p_{hi} = overall sampling probability for households selected for the NHFS in the i-th sample EA in strata h,

n_h = number of sample EAs selected in strata h for the NHFS,

M_{hi} = total number of households in the i-th sample EA in strata h from the 2008 PHC frame,

M_h = total number of households in strata h from 2008 PHC frame ,

m_{hi} = 12 = number of sample households selected for the NHFS in the i-th sample EA in strata h, and

M'_{hi} = total number of households in the new listing for the i-th sample EA in strata h.

The basic weight for the NHFS sample households is the inverse of this probability of selection, expressed as follows:

$$W_{hi} = \frac{1}{p_{hi}} = \frac{M_h \times M'_{hi}}{n_h \times M_{hi} \times m_{hi}}, \text{ where:}$$

W_{hi} = basic weight for the NHFS sample households in the i-th sample EA in strata h.

In order to account for the outdated nature of the 2008 PHC frame and the erroneous distance-based stratification, the basic weights were then adjusted for: (i) cluster level populations, and (ii) population growth since 2008. First, the EAs were categorized by cluster (as defined in Section 2.6). Then, a weight adjustment factor was computed based on the Liberia 2016 HIES population estimates, assuming the cluster level population growth rate from the 2008 PHC to the 2016 HIES population estimates.¹⁰ The weight adjustment factor based on the 2016 HIES population estimates can be expressed as follows:

$$A_c = \frac{P_{HIESc}}{\sum_{ic} \sum_j W'_{hi} \times p_{cij}}, \text{ where:}$$

A_c = adjustment factor for the weights of NHFS sample households in cluster c,

P_{HIESc} = total population for cluster c based on the 2016 HIES estimates

W'_{hi} = basic design weight for the sample households in the i-th sample EA in strata h, and

p_{cij} = number of persons in the j-th sample household of the i-th sample EA in cluster c.

The denominator of the adjustment factor A_h is the estimated total population in cluster c from the NHFS data using the basic design weights. The design weights for all the sample households within a district were multiplied by the corresponding adjustment factor for the district to obtain the final adjusted weights, as follows:

$$W_{Aci} = W_{hi} \times A_c, \text{ where:}$$

W_{Aci} = final adjusted weight for the sample households in the i-th sample EA in cluster c.

The resulting weight adjustment factors for each cluster are: North cluster – 1.294602; Central cluster - 1.294602; Eastern cluster – 1.584155.

¹⁰ The population growth rate is assumed to be the same for all EAs in the cluster, irrespective of the distance from forest. This is a necessary assumption as population estimates are not available at the EA level.

7. INCOME AGGREGATE

The income aggregate for the NHFS follows the Rural Livelihoods Information System (RuLIS) methodology which results in a comprehensive measure of household income.¹¹ This includes information on different sources of income, such as wage employment (both agricultural and non-agricultural, forest and non-forest), self-employment, crop, forestry and mining production, transfers, and other sources of income, such as non-labor earnings. Note that data on income from livestock and fisheries are not collected in the NHFS questionnaire explicitly. In order to accommodate the extensive forestry data collected in the NHFS, the traditional RuLIS methodology was adapted slightly to provide more disaggregated forestry variables. All income variables are reported in Liberian Dollars, at the time of the survey.

Variable *totincome* presents the total net household income, while *pcinc* presents the per capita income. The main components of total income are described below.

Employee income – wages

Wages are employees' compensations received either in cash or kind from primary and secondary jobs during the last 12 months. Real wages are collected and disaggregated by sector, as follows: (i) agriculture, forestry & fishing (*wge1*); (ii) mining and quarrying, manufacturing & construction (*wge2*); (iii) services (*wge3*); and (iv) other industries (*wge4*).

Income from self-employment

Income from self-employment is a *net figure* that counts cash earnings from sales of produced goods or provided services and non-durable expenditures on raw materials (i.e. pesticides, fertilizer, vaccines, etc.), wages/salary, and other operational costs (i.e. rent, fuel, kerosene, electricity, transport, and marketing) over the last 12 months.

This income component is collected and disaggregated by sector, as follows: (i) agricultural enterprises; (ii) construction-extraction related enterprises, such as mining, handicraft manufacturing, and construction; (iii) services-related enterprises, such as wholesale and retail trade, transportation and storage, accommodation and food service activities, financial and insurance activities, and education activities; (iv) forest-related enterprises, such as trade of forestry products, business of organizing/skilled labor in forest-related activity (i.e. ranger service), transport of forest products, and ecotourism business.

Revenues and expenses are annualized based on reported number of months in operation in last 12 months. Variable *selfemp* presents the total income from self-employment activities. See variables *self1*, *self2*, *self3*, *self4* for sector specific self-employment incomes.

¹¹ For resources related to the RuLIS methodology, see: <http://www.fao.org/in-action/rural-livelihoods-dataset-rulis/resources/en/>

Income from on-farm activities – crop production

Household income from on-farm activities consists of incomes from crop production activities and other farm income. This component refers to net income, which is defined as an operating surplus (i.e. revenues minus operating costs).

Total net income from crop production (*crop*) is equal to the monetary value of the total harvested quantity minus operating costs. In general, the value of the total harvested quantity is the value of all crop quantities used for different purposes, including those marketed (i.e. crop sold, bartered or provided as payments in kind to hired labor) and consumed (i.e. value of crop consumed by the household or retained for use in future production). Operating costs comprise all variable costs (i.e. payments in cash, all kinds of agricultural inputs, such as fertilizer and seeds) and fixed costs (i.e. land rent).

Information on crop disposition is collected in non-standard units of measurement and in unit prices for harvested, sold, consumed, in-kind, gift, animal feed and saved crop. The monetary value of the total harvested quantity is imputed using unit sales prices that are considered to be more consistent with the current market prices than the reported harvested unit prices. Unit sales prices are computed using the median unit prices of every crop for the closest possible geographic and sample entities (i.e. enumeration area, clan, district, county, or national).

Revenues (+)	Costs (-)
Crop sold	Inputs paid in cash
Crop used for own consumption	Land rent
Crop used as feed	Crop saved for seed
Crop stored	Crop used for paying in-kind labor
Crop given as gift	Crop used for paying inputs
Crop saved for seed	
Crop used for paying in-kind labor	
Crop used for paying inputs	

Income from forestry

The methodology used for imputing income from forestry follows the general principles presented in the crop production section with some additional sources of income from enterprise and non-enterprise forest-related activities.

Income from forestry production is equal to the monetary value of the total forestry production value minus operating costs. The monetary value of the total forestry production is the value of all harvested and processed forestry quantities sold, consumed or retained for future uses. Operating costs comprise transport and marketing costs, costs of purchased and own inputs, hired labor and any costs of renting land/collection fees both for harvested and processed forestry products.

Information on forestry products is collected in non-standard units of measurement for all harvested and processed forestry quantities produced, sold and consumed. Unit prices are available in the NHFS questionnaire for all harvested and processed forestry quantities sold and consumed,

but not produced. Thus, the monetary value of the total forestry production is imputed using unit sales prices. As for the crop production, unit sales prices are computed using the median unit values of every forestry product for the closest possible geographic and sample entities (i.e. enumeration area, clan, district, county, or national).

Finally, net income from non-enterprise and enterprise forest-related activities are computed separately. *Net income from non-enterprise* forest-related activities is equal to the monetary value of the total forestry production value plus Payments for Environmental Services (PES). *Net income from enterprise* forest-related activities imputes revenues and expenses from forestry-related enterprise activities over the last 12 months (see ‘Self-employment income’). *Net Income from all forest activities* includes both net income from non-enterprise and enterprise forest related activities.

Revenues (+)	Costs (-)
Income from forestry production (harvested and processed forestry products)	Inputs costs (land rent, fertilizers, hired labor, etc.)
Income from non-enterprise forest activities, including PES programs	Transport costs
	Marketing costs

Note: Income from forest-related enterprise activities is included in the Self-employment Income. The value of production from mining/mineral collection and processing is imputed separately from forestry incomes

The variable *forestinc_1* includes income from the collection and processing of forest and wild products and income from PES.

The variable *forestinc_2* includes income from the collection and processing of forest and wild products, income from PES, and income from forest-related enterprises.

Income from mining/mineral activities

Net income from mining activities (variable *mininginc*) is equal to the monetary value of the total mining products collected and processed minus the operating costs. The monetary value is the value of the total quantities sold, consumed or retained of all mining products, such as gold and diamonds. Operating costs comprise transport/marketing costs, costs of purchased and own inputs, hired labor and any costs of renting land/collection fees both for collected and processed mining products.

As for forest and crop products, information on mining products is collected in non-standard units of measurement for all collected and processed mining quantities produced, sold and consumed. Unit prices are available in the NHFS questionnaire for all collected and processed mining quantities sold and consumed, but not produced. Thus, the monetary value of the total mining production is imputed using unit sales prices. As for forestry and crop production, unit sales prices are computed using the median unit values of every mining product for the closest possible geographic and sample entities (i.e. enumeration area, clan, district, county, or national).

Revenues (+)	Costs (-)
Income from mining production (collected and processed mining products)	Inputs costs (land rent, fertilizers, hired labor, etc.) Transport costs Marketing costs

Income from transfers

This income component accounts for private and public transfers received by the household in the last 12 months. The definition of transfer income and the construction of its components are based on the RIGA methodology (FAO) according to which private transfers refer to private cash, food, in-kind transfers from international and national sources (i.e. incoming remittances). According to the same methodology, public transfers are divided into: (i) state-funded pensions and social security; (ii) and social benefits, which include welfare support, maternity benefits, and educational transfers. For total transfers, see variable *transfer*.

Other sources of income

Other sources of income include non-labor income components that do not fall into the previous categories described above. The sources of income included in the figure include rent derived from land, real estate and owned assets and financial revenues from savings and interest rates. See variable *otherinc*.

Income shares

The calculation of the shares of income from various sources (*share_**) may show the presence of negative values in productive activities, where income is measured as a difference between revenues and costs. Negative values in total income imply a net loss for the household. There may be negative values in the income shares associated with particular income sources where households have a loss from a given source. Negative values, and those exceeding one, are retained in the dataset presented here. However, according to RuLIS, the observations with income shares less than zero and greater than one should be dropped when analyzing mean income shares for the population.

Socio-demographic variables

In addition to the income variables, the income aggregate file includes some constructed variables on household socio-demographics, including household size, average educational attainment, age of the household head, etc.

REFERENCES

- FAO, CIFOR, IFRI and World Bank. 2016. National socioeconomic surveys in forestry: guidance and survey modules for measuring the multiple roles of forests in household welfare and livelihoods, by R.K. Bakkegaard, A. Agrawal, I. Animon, N. Hogarth, D. Miller, L. Persha, E. Rametsteiner, S. Wunder and A. Zezza. FAO Forestry Paper No. 179. Food and Agriculture Organization of the United Nations, Center for International Forestry Research, International Forestry Resources and Institutions Research Network, and World Bank.
- LISGIS (Liberia Institute of Statistics and Geo-information Services). 2016. *Household Income and Expenditure Survey (HIES) 2016*. Monrovia, Liberia: LISGIS
<http://microdata.worldbank.org/index.php/catalog/2986>
- Metria and Geoville. 2019. (Authors: Rosengren, Mats, Jürgen Weichselbaum, and Norman Kiesslich.) *Update of the Liberia Land Cover and Forest Map*. GeoVille Information Systems and Data Processing GmbH, Innsbruck, Austria.

ANNEX I. SAMPLE SIZE CALCULATIONS

When populations are heterogeneous, *ceteris paribus*, a larger sample is more desirable than a smaller sample, as it would give statistically more reliable results. The principal objective of stratification is to reduce sampling errors. In a stratified sample, the sampling errors depend upon the population variance existing within the strata but not between the strata. Thus, it helps to create strata with low internal variability. The initial choice of the 3 strata, based on distance to the forest, is an attempt to ensure that households, within each stratum, are fairly homogeneous.

A common measure of precision for estimating an indicator (or variable of interest) is its Relative Standard Error (RSE) (the standard error of the indicator divided by its estimated value). Then the required sample size “n”, for a desired RSE “ α ”, an indicator value, “P” (the major variable of interest, defined as the percentage of households’ dependent on forests), and a design effect¹² “Deft”, is given by:

$$n = \text{Deft}^2 \times ((1/P - 1) / \alpha^2) \quad (1)$$

As shown in Table 7 below, to calculate the sample size for each of the 3 strata, we have chosen a Deft value of 2.5 for each, as no independent information on this parameter is available for Liberia.¹³

The main variable of interest, “P”, is the proportion of households dependent on forest products of one sort or another. In addition to wood, other products that households depend upon include fuelwood and charcoal, medicinal plants, bushmeat and non-timber forest products (such as wild fruits and nuts, bush pepper, grains of paradise, bitter cola, etc.). Based on the available literature for Liberia, household dependency on one or more of these products can be as high as 80-90%, especially for households living close to forests. For purposes of estimating the sample sizes in each of the 3 strata, the P values are taken to be 0.8, 0.7 and 0.6 (based on the available literature, which indicates a high level of dependency for households in proximity to the forest, and using our judgement to assume lower dependency ratios, the further households are to the forest margins). The α values, for error tolerance are taken as 0.04, 0.05 and 0.06, for the 3 strata. Using these parameters yields the samples sizes of 977, 1071 and 1157, respectively, with a total sample size of 3205 households. For each stratum, the expected standard errors are less than 4% in each case (keeping standard errors below 5%, is generally considered desirable).

¹² Deft is a measure of the efficiency of cluster sampling compared to a direct simple random sampling of individual households. A simple random sampling is considered the most efficient approach and the Deft value adjusts for the loss of efficiency, by increasing the sample size.

¹³ In the absence of any information, it is frequently recommended that a Deft value of 2 be used. We are being more risk averse and choosing a higher value of 2.5. Roughly speaking, we are saying that the sample design for this survey is only 40% as efficient (as measured by the standard error of the indicator) as compared to a simple random sample, and so we increase the sample size by the square of 2.5, to get the same level of efficiency.

Table 7. Sample Size Estimates

Strata	Deft	P variable	Desired RSE	N	Expected Std. Error
S1	2.5	0.8	0.04	976.6	0.032
S2	2.5	0.7	0.05	1071.4	0.035
S3	2.5	0.6	0.06	1157.4	0.036
Total				3205.4	

ANNEX II. DATA FILE STRUCTURE

Table 8. Structure of HH Data Files

File Name	Module Name	Level of Analysis	Identification Variable(s)
sect1_public	Section 1: Household identification & Survey staff details	Household	hhid
sect2_public	Section 2: Household member roster	Individual	hhid member_id
sect3_public	Section 3: Forest resource base	Household	hhid
sect4_public	Section 4: Forest benefits	Forest service	hhid serviceID
sect5_public	Section 5: Forests and health	Household	hhid
sect6_public	Section 6: Forests and energy	Forest product	hhid prod_id
sect7_filter_public	Section 7: Income from forests (collected products)	Household	hhid
sect7_public	Section 7: Income from forests (collected products)	Forest product	hhid prod_id
sect8_filter_public	Section 8: Income from forests (processed products)	Household	hhid
sect8_public	Section 8: Income from forests (processed products)	Forest product	hhid prod_id
sect9_public	Section 9: Labor income	Individual	hhid member_id
sect10_filter_public	Section 10: Land parcels	Household	hhid
sect10_public	Section 10: Land parcels	Parcel	hhid parcel_id
sect11_public	Section 11: Crop disposition	Crop	hhid crop_id
sect12_filter_public	Section 12: Household non-farm enterprises	Household	hhid
sect12_public	Section 12: Household non-farm enterprises	Household Enterprise	hhid enterprise_id
sect13_filter_public	Section 13: Other forest income including PES	Household	hhid
sect13_public	Section 13: Other forest income including PES	Program	hhid practice_id
sect14_public	Section 14: Other non-labor income	Source of income	hhid source_id
sect15_public	Section 15: Household assets	Assets	hhid item_id
sect16a_public	Section 16: Food security	Household	hhid
sect16b_public	Section 16: Food security	Forest product	hhid prod_id
sect17_public	Section 17: Shocks and crises	Shock	hhid shock_id
sect18_public	Section 18: Household access to forests	Household	hhid
sect19_public	Section 19: Forests and construction	Household	hhid

sect20_public	Section 20: Forest clearance	Household	hhid
sect21_public	Section 21: Household recontact information	Household	hhid

Table 9. Structure of Community Data Files

File Name	Module Name	Level of Analysis	Identification Variable(s)
comm_sect1_public	Section 1: Cover	Community	ea_unique
comm_sect2_public	Section 2: Roster of Informants	Informant	ea_unique number
comm_sect3_public	Section 3: Forest Roster	Forest	ea_unique forest_id
comm_sect4_public	Section 4: Seasonal Calendar	Forest Product	ea_unique prod_id
comm_sect5_public	Section 5: Community Forest Status	Community	ea_unique
comm_sect6a_public	Section 6a: Most Important Forest and Wild Products (Cash)	Forest Product	ea_unique prod_id
comm_sect6b_public	Section 6b: Most Important Forest and Wild Products (Subsistence)	Forest Product	ea_unique prod_id
comm_sect7_public	Section 7: Units and Pricing	Forest Product	ea_unique prod_id
comm_sect8a_public	Section 8a: Community Participation	Community Program	ea_unique practice_id
comm_sect8b_public	Section 8b: Support for forest-related activities	Community Program	ea_unique support_id
comm_sect9a_public	Section 9a: Gender Activities (Male respondents)	Community	ea_unique
comm_sect9b_public	Section 9b: Gender Activities (Female respondents)	Community	ea_unique

Table 10. Structure of Supplemental Data Files

File Name	Description	Level of Analysis	Identification Variable(s)
Liberia_NHFS_HH_Income.dta	Income Aggregate	Household	hhid
Liberia_NHFS_HH_DistanceToForest.dta	Distance from HH to nearest forest, by level of canopy cover	Household	hhid