



SUDAN



SUDAN STRATEGY SUPPORT PROGRAM WORKING PAPER 18

APRIL 2024

Sudan Rural Household Survey 2023

Sampling and implementation procedures for the first round

Oliver K. Kirui, Mosab Ahmed, Alemayehu Seyoum Taffesse, Hala Abushama, and Khalid Siddig

ABSTRACT

This paper describes the data collection methodology for the 2023 Sudan Rural Household Survey (SRHS), the first nationwide survey of rural households in Sudan following the eruption of violent conflict in April 2023. The SRHS, which was conducted by the International Food Policy Research Institute in collaboration with the United Nations Development Program and the World Food Program, represents a significant effort to understand the dynamics of rural households in Sudan during a period of profound national conflict. The survey focuses on several critical themes central to household welfare, including food security, coping strategies, employment and income, livelihoods, assets, market access, and exposure to shocks. Analysis of the survey data can provide comprehensive insights into the resilience of and challenges faced by rural communities in Sudan. Computer-assisted telephone interview (CATI) methods were used to navigate the challenges of data collection in a conflict setting. The use of CATI demonstrated the adaptability and potential of innovative research methodologies in conflict-affected regions. The survey, the first round in a planned panel survey program, aims to establish a baseline of rural life, economic activities, and food security in Sudan. Panel survey data offers invaluable information to counter several analytical biases inherent to the conclusions drawn from other data structures. Subsequent rounds of the SRHS are planned for the second and fourth quarters of 2024.

Key words: Sudan, rural household survey, conflict, computer-assisted telephone interview.

CONTENTS

- Abstract.....i
- 1. Introduction1
- 2. Sample Size Determination and Selection.....2
 - 2.1 Sample size for the telephone survey2
 - 2.2 Sample selection4
- 3. Questionnaire and Data.....6
 - 3.1 Modules and key variables6
 - 3.2 Questionnaire review.....8
- 4. Data Collection and Training8
 - 4.1 Call center preparation and operators’ training9
 - 4.2 Survey implementation: data entry, transmission, and delivery10
 - 4.3 Quality assurance and internal control mechanisms10
 - 4.4 Privacy, data security, and ethics11
- 5. Implementation Challenges12
- Annexes13
 - Annex 1—Household survey sample design factors13
 - Annex 2—Sample size determination for the 2023 Sudan Rural Household Survey...14
- About the Authors.....16
- Acknowledgments16

TABLES

- Table 2.1 Estimated minimum detectable effect (MDE) for different sample sizes based on poverty incidence2
- Table 2.2 Share of Sudan’s population and number of sample households by state3
- Table 2.3: Distribution of telephone numbers obtained from databases of the World Food Programme and GeoPoll, by state.....4
- Table 2.4: Calls attempted during data collection6
- Table 4.1 Implementation calendar for the 2023 Sudan Rural Household Survey8

1. INTRODUCTION

The 2023 Sudan Rural Household Survey (SRHS), conducted by the International Food Policy Research Institute (IFPRI) in collaboration with the United Nations Development Program (UNDP) and the World Food Program (WFP), is the first nationally representative household survey conducted following the start of the current conflict in April 2023. The survey was based on a sample of 4,504 households who were interviewed using computer-assisted telephone interview (CATI) methods. The survey's findings and outputs are poised to contribute significantly to the discourse on rural development, resilience building, and humanitarian assistance in Sudan and elsewhere.

The 2023 SRHS emerges at a critical juncture in Sudan's socioeconomic landscape. It aims to illuminate the complexities and resilience of rural households amid a backdrop of significant national upheaval. This survey seeks to establish a baseline for understanding the multifaceted dimensions of rural life, economic activity, and food security within the country during the conflict.

The strategic decision to employ a longitudinal panel study approach underscores a commitment to trace the dynamic evolution of rural households over time, particularly their changes in poverty, food security, and livelihoods, considering the evolving conflict in Sudan. The survey's inclusive scope, covering rural households and, in future phases, urban households and micro, small, and medium enterprises, offers a holistic strategy for assessing economic conditions and societal well-being in Sudan under conflict. The nuanced insights obtained will assist in formulating effective policy measures to bolster the resilience and sustainable development of communities across Sudan.

However, the outbreak of conflict in April 2023 required a shift in data collection strategies. Conventional in-person interview approaches, as used in earlier household surveys and censuses, while rich as a result of the in-depth engagement required, are untenable and impractical during conflict. The adoption of CATI for the SRHS not only exemplified adaptability under crisis but also demonstrated the potential of innovative data collection techniques to surmount research barriers in areas afflicted by conflict. This methodological pivot ensured the continuity of research activities, providing timely insights for both immediate and strategic planning and the design of interventions.

The 2023 SRHS exemplifies the transformative impact of collaborative research and innovation in surmounting conflict-related challenges. The insights derived from the analysis of the SRHS data will influence strategies and policies aimed at reducing poverty, improving food security, and enhancing economic resilience among Sudan's rural populations. This ongoing research initiative, thus, enriches our comprehension of rural Sudan amid adversity and serves as a model for informed decision-making in crisis-affected environments.

2. SAMPLE SIZE DETERMINATION AND SELECTION

2.1 Sample size for the telephone survey

The primary objective of the SRHS is to track changes in indicators of household economic productivity, resilience, coping strategies, and well-being on a national scale, covering all states as security conditions allow.

Assessing the size of the minimum detectable effect (MDE) for a range of sample sizes was used as a practical approach to guide the choice of the size of the sample for the survey. Data from the most recent representative survey for Sudan, the 2022 Sudan Labor Market Panel Survey (SLMPS), would have been optimal for this assessment, as it provides the most recent pre-conflict data on labor market dynamics, employment patterns, incomes, prices, and poverty rates. The SLMPS was implemented by the Central Bureau of Statistics and the Economic Research Forum using in-person interviewing and was the first comprehensive household-level data produced for Sudan since the 2014/15 Household Survey and the 2008 National Census. However, while data analysis of the SLMPS was completed by August 2023, when the 2023 SRHS was being designed, the SRHS survey team was unable to obtain access to the data. Rather, data from the earlier 2014 Sudan Multiple Indicator Cluster Survey (MICS) was used to determine the appropriate size for the SRHS sample. MICS had also been used in this way for the design of the 2022 SLMPS sample.

We evaluated sample sizes ranging from 1,200 to 4,500 households, aiming for 80 percent power and a 5 percent significance level, and examined their corresponding minimum detectable effect (MDE) in measuring poverty incidence (Table 2.1).¹ Increasing the sample size offers diminishing returns in terms of MDE improvement. Nevertheless, considering the prevailing conflict-related uncertainties, the telephone-based data collection method, and the somewhat limited scope for changing the sampling frame, it was judged wise to choose the largest feasible sample size. (The sample size determination is described in more detail in Annex 1.) Therefore, we chose a sample size of 4,500.

Table 2.1 Estimated minimum detectable effect (MDE) for different sample sizes based on poverty incidence

Sample size	1,200	2,000	3,000	4,000	4,500
Estimated MDE, percentage points	4.04	3.13	2.56	2.21	2.09

Source: Authors' calculation
Note: Assessment based on 80 percent power and a 5 percent significance level,

Use of the telephone-based CATI method to implement the survey rather than in-person interviewing requires, however, several notable adjustments:

- ▶ The sampling frame was not a comprehensive list of households but a compilation of telephone numbers or individuals associated with a telephone number. The telephone

¹ In addition to sample size, power, and significance, this pared down description of the exercise to determine an optimal survey sample size uses a starting poverty headcount rate of 50 percent to compute the MDEs. This poverty rate is also used in the more detailed power calculations described in the Annex. Note, however, that the analysis described in the Annex accounts for other dimensions of the survey and its context, such as the design effect and attrition, and includes those factors in determining sample sizes for given MDEs.

number databases used covered all states in Sudan and were drawn from national surveys carried out by WFP between 2018 and 2022.

- ▶ Using CATI survey methods, the predictability of the final sample size is significantly lower than for in-person surveys due to factors such as no longer active numbers in the telephone number databases and network and connectivity challenges.
- ▶ Allocating the sample evenly across states required a logical criterion. We adopted the proportional distribution of states within Sudan’s total population as the primary guiding principle for allocating the sample across states.

These adjustments ensure a methodologically sound basis for conducting a survey that aims to produce accurate and reliable state and national-level inferences despite the inherent challenges of telephone-based data collection. The final sample for the SRHS by state is presented in Table 2.2.

Table 2.2 Share of Sudan’s population and number of sample households by state

State	Sudan’s population living in the state in 2014, percent	Sample households in the state number	percent of total
Khartoum	13.8	621	13.8
Central Darfur	1.8	81	1.8
East Darfur	3.0	135	3.0
North Darfur	7.4	333	7.4
South Darfur	7.6	342	7.6
West Darfur	3.3	149	3.3
North Kordofan	6.7	302	6.7
South Kordofan	2.8	126	2.8
West Kordofan	6.0	270	6.0
Sennar	3.9	176	3.9
Aj Jazirah	15.6	702	15.6
Blue Nile	3.9	176	3.9
White Nile	5.2	234	5.2
Northern	2.5	113	2.5
River Nile	4.0	180	4.0
Gedaref	5.1	230	5.1
Kassala	4.3	194	4.3
Red Sea	3.1	140	3.1
Total	100.0	4,504	100.0

Source: Authors’ compilation.

Note: Sample households are those for which telephone numbers were obtained. The full sample was achieved after an extension of the survey period and repeated calls.

This survey leveraged the WFP telephone number database, consisting of 29,724 numbers. The database is linked to WFP food security assessment surveys from 2018 up to and including 2022, which were nationally representative of the general population across Sudan. This linkage to earlier surveys enables us to utilize pre-conflict data on respondents in our analyses where warranted, especially from 2022. This additional information on the survey households over time provides a distinct advantage for analysis and makes the use of such telephone number databases from earlier surveys a preferred choice for sampling frames in telephone surveys.

However, the availability of telephone numbers in the WFP database was limited in Northern, River Nile, and South Kordofan states. In addition, for North Kordofan, Red Sea, Khartoum, Blue Nile, and Gezira states, achieving the desired sample size with the existing database required exceptionally high call completion rates of between 35 and 93 percent. Therefore, the WFP database was augmented with a telephone number database from GeoPoll, one of the private firms contracted to carry out the CATI-based administration of the 2023 SRHS. Particularly in states where the WFP database had insufficient numbers to attain the desired sample size, the GeoPoll telephone database, consisting of 24,824 numbers nationally, was used as a supplementary primary sampling frame. This approach ensured a more comprehensive and representative sampling strategy. The distribution of the two databases by state is presented in Table 2.3.

Table 2.3: Distribution of telephone numbers obtained from databases of the World Food Programme and GeoPoll, by state

State	World Food Programme database	GeoPoll database	Total
Khartoum	4,864	4,996	9,860
Central Darfur	679	553	1,232
East Darfur	1,002	495	1,497
North Darfur	1,603	1,644	3,247
South Darfur	2,022	1,278	3,300
West Darfur	868	825	1,693
North Kordofan	4,000	1,802	5,802
South Kordofan	698	1,253	1,951
West Kordofan	1,915	1,561	3,476
Sennar	1,070	348	1,418
Aj Jazirah	1,440	597	2,037
Blue Nile	646	1,380	2,026
White Nile	1,415	2,316	3,731
Northern	348	986	1,334
River Nile	543	1,645	2,188
Gedaref	1,635	222	1,857
Kassala	1,518	1,793	3,311
Red Sea	3,458	1,130	4,588
Total	29,724	24,824	54,548

Source: Authors' calculation.

2.2 Sample selection

Phase one of the SRHS used a stratified random sampling approach to select the sample households. The state was the stratification variable. Within each state, respondents were selected randomly from the telephone databases with an equal probability of selection, ensuring the representativeness of our sample at the state level.

Sample selection from the two telephone databases, i.e., the sampling frames, proceeded as follows:

1. The sample selection process started with the WFP telephone number database:
 - a. The telephone numbers were organized by state and locality, utilizing matched information from previous WFP surveys to do so.

- b. The required sample size for each state was calculated and divided by the number of localities within a state to determine the target sample size per locality (denoted as 'y').
- c. Beginning with the least populous locality within a state, an assessment was made of whether the available telephone numbers in each locality in the telephone number database were sufficient for selecting the desired sample, 'y', for that locality.
 - i. If the count exceeded 'y', telephone numbers were randomly selected to reach the desired sample size, employing a random number generator for fairness.
 - ii. If the count fell below 'y', then all available telephone numbers in the WFP database were used to compile the sample. When all those numbers had been tried, potential sample households were contacted by randomly selecting telephone numbers from the GeoPoll database until 'y' was met.
- d. If both databases proved insufficient for achieving 'y' in a locality, we allocated the shortfall to the sample of the next larger locality within the state based on population size, repeating steps (c.) and (d.) as needed.

The research team supplied GeoPoll, one of the firms contracted to implement the CATI survey, with 29,724 mobile telephone numbers from the WFP database in two phases—14,124 numbers were supplied at the start of interviewing, with an additional 15,600 numbers provided later in survey implementation. In addition, the WFP database was compared to the GeoPoll database, and entries in the GeoPoll database that duplicated those in the WFP database were eliminated. Prior to assigning the numbers to the call center manager and enumerators, the mobile telephone numbers were randomized to ensure the survey's representativeness.

2. During the data collection phase:

- a. Sample integrity was ensured by auditing call dispositions to assess response rates and develop improvement strategies.
- b. Respondent contact attempts were limited to four, with further attempts classifying the number as non-responsive, as detailed in Table 2.4.
- c. The consent of the respondent to be interviewed was recorded for potential review and clarification of responses.

Table 2.4: Calls attempted during data collection

Call Status	Count	Percent
Call attempts	74,473	100.0
Of the call attempts:		
No response	62,136	83.4
Reached	12,337	16.6
Of those reached:		
Refused	5,474	44.8
Agreed	6,739	55.2
Of those who agreed to be interviewed		
Completed	4,504	66.8
Dropped-off	519	7.7
Ineligible	1,716	25.5

Source: Authors' calculation.

Note: Up to three additional follow-up calls were made to numbers for which no response was initially obtained.

3. Post-interview incentives

Respondents who completed the interview received an incentive of standard airtime credit within two days of survey completion, affirming their valuable participation. With the advice of the survey company, GeoPoll, and based on their previous experience in conducting CATI surveys in Sudan, airtime credit equivalent to about 2.00 US dollars, was considered sufficient compensation for respondents at the end of the interview as an appreciation for the time they spent being interviewed.

3. QUESTIONNAIRE AND DATA

The first round of the SRHS used a detailed questionnaire designed to capture a wide array of data crucial for understanding the economic status, food security, living conditions, coping mechanisms, and overall well-being of rural Sudanese households amid the country's persistent conflict. The questionnaire includes a series of modules, each on different facets of rural life. These encompass demographic information, livelihood approaches, food security levels, market accessibility, and the receipt of aid.

The questionnaire administration started with introductory questions to verify respondent eligibility, subsequently unfolding through a structured series of modules that investigate a variety of thematic areas. This structured approach ensured a comprehensive understanding of the multifaceted challenges faced by rural communities. This modular approach enables a holistic understanding of the complex interplay between demographics, economics, security, and well-being to yield valuable insights for policymaking and intervention strategies aimed at enhancing the resilience and well-being of rural communities in Sudan.

3.1 Modules and key variables

Module A: Introduction and Informed Consent: This initial module serves as an entry point for the survey where the interviewers present themselves and explain the study's objectives. The eligibility of participants is verified. Informed consent from the respondent is also obtained in this module, ensuring that the responding individual understands that their

responses will be held confidentially and emphasizing their right to withdraw from the survey administration process at any time.

Module B: Demographic and Geographic Information: This detailed demographic module collects information on age, sex, and relationship to the household head for all household members, creating a comprehensive demographic profile of each respondent household. This section also includes geographic questions to document the current and pre-conflict residences of respondents to provide data for regional analysis and the development of targeted policies.

Module C: Household Composition: This module collects additional information on household composition, including the marital status of the household head and any recent relocations of household members due to conflict. This information is needed to better understand recent household dynamics and the impacts of displacement.

Module D: Income Earners and Income Sources/Employment: The focus of this module is on economic activities, exploring the main income-generating activities of the household and associated challenges for each. The survey captures pre- and post-conflict employment, income levels, agricultural practices, and the immediate effects of the conflict on respondents' ability to maintain their livelihoods.

Module E: Food Consumption Patterns: Respondents detail their household's food consumption, offering insight into the diversity in household diets and the frequency of intake of various food groups. This data is used for assessing nutritional status, while also highlighting some dimensions of the relative food security of respondent households.

Module F: Food Security Assessment: This module gathers information on experiences of food insufficiency, anxiety about future food provision, and any associated coping behaviors. This module is used to gauge the prevalence and severity of food insecurity among the surveyed population.

Module G: Coping Strategies Index: Information on household responses to food and income shortages is collected in this module. It examines whether families have had to sell assets, reduce expenses, or adopt other strategies to cope with scarcity.

Module H: Household and Agricultural Assets: This module collects information on dwelling types, dwelling ownership, and household assets. This information is important for indicating the socioeconomic status and quality of life of a household. Information is also obtained on sources of basic services for the household, such as water and electricity, as well as information on the agricultural lands owned by the household, if any.

Module I: Market Access and Disruptions: In this module, respondents reflect on their access to local markets, any obstacles encountered in selling or purchasing goods, and the comparative ease or difficulty of market transactions before and after the eruption of the conflict. Analysis of the data obtained using this module will help in understanding the economic impact of the conflict on local trade and market functionality.

Module J: Shocks: This module captures data on the various shocks, both natural and violence-related shocks, that households have faced. It also gathers information on the respondents' perceptions of physical security.

Module K: Assistance and Support: In this module, interviewees report on the assistance that their household has received from various sources and any changes in sources or levels of assistance due to the conflict. Data from this module can guide the design and targeting of humanitarian and development aid to ensure it meets recipients’ most pressing needs.

Module L: Closing and Future Engagement: The final module wraps up the interview, confirming the best means for future contact with the respondent, setting the stage for potential follow-up interviews. This ensures the continuity of the survey effort and the integrity of longitudinal data collection.

3.2 Questionnaire review

The development of the survey questionnaire followed a meticulous and collaborative process, where the refinement of the questionnaire involved a comprehensive review of the changes made to the original draft, ensuring that the revised questions were coherent and effectively captured the intended data. Following the approval of the English version of the questionnaire, the translation process commenced. The survey was translated into the dialect of Arabic spoken in Sudan, which was the language used to administer the survey. The research team closely scrutinized the translation to maintain the accuracy of the survey questions.

Questionnaire scripting took place after a thorough review of the translations. Post scripting of the questionnaire, the script was tested via pilot administration of the questionnaire using the CATI application to ensure the correct flow of the scripting was followed.² In preparation for data collection, the call center enumerators were also given access to a survey script, allowing them to acquaint themselves with the questionnaire’s flow and content.

The careful orchestration of these steps ensured that the survey instrument was robust, culturally relevant, and ready for deployment in the field. This comprehensive approach to survey design and testing underpins the integrity and validity of the data collection process, which is fundamental for research endeavors aiming to inform policy and intervention strategies.

4. DATA COLLECTION AND TRAINING

The calendar for the implementation of the 2023 SRHS is presented in Table 4.1. This section of the report provides further details on several of the implementation steps.

Table 4.1 Implementation calendar for the 2023 Sudan Rural Household Survey

Task	Description	Dates	Duration (in days)
Training of the trainers for household survey enumerators	GeoPoll led the training of trainers for the household survey with the support of IFPRI. The CATI manager for the survey and enumerator supervisors participated in the training, which was conducted in English.	25 October 2023	1
Questionnaire and script updates following training of trainers	GeoPoll, UNDP, and IFPRI staff revised the questionnaire using insights gained from the training of trainers sessions.	26-30 October	3

² Sixty-eight interviews were conducted during the piloting phase.

Task	Description	Dates	Duration (in days)
CATI Enumerators Training	The GeoPoll CATI manager led the training of enumerators, focusing on project objectives, survey methodology, and the household survey questionnaire.	1-2 November	2
Pilot data collection	Following their training, the enumerators conducted pilot interviews to build their skills and to test the questionnaire modules. IFPRI researchers reviewed the data collected through the pilot interviews, modifying the questionnaire as needed.	3-6 November	4
Debrief after the pilot interviews	GeoPoll and IFPRI conducted a debrief session with enumerators after the pilot interviews to discuss any changes to the questionnaire or survey design	7-8 November	2
Daily data collection after approval of pilot data	GeoPoll commenced CATI data collection for the SRHS after finalizing updates to the questionnaire from IFPRI and UNDP.	9 November 2023 to 5 January 2024	58
Translations and coding of open-ended data	After each interview, GeoPoll translated any open-ended responses into English and coded the data.	7-10 January 2024	4
Final dataset reviewed by GeoPoll	GeoPoll incorporated the coded responses into the final dataset.	11-12 January	2
Cleaned final dataset submitted to IFPRI	GeoPoll delivered the cleaned final dataset to IFPRI in Excel, SPSS, and Stata formats.	13 January	1

Source: Authors' compilation

Note: GeoPoll was one of the firms contracted to handle data collection for the 2023 SRHS.

4.1 Call center preparation and operators' training

Based on the determination of the desired sample size for the survey of 4,504 households, a team of 34 enumerators under two supervisors was recruited to handle data collection. The enumerators recruited were experienced in administering food security-related surveys, including using telephonic CATI methods.

A training-of-trainers approach was used in training the survey implementation staff. The IFPRI research team and GeoPoll management and technical staff trained in English the CATI managers and enumerator supervisors on the objectives and content of the survey. After that, the managers and supervisors trained the enumerators in Sudanese Arabic under the mentorship of the research team. Training of enumerators was conducted virtually over three business days.

During the training, the purpose of the survey was explained, 70 pilot interviews were conducted in Sudanese Arabic, and any questions that needed to be rephrased were updated. A debriefing session was conducted with the enumerators after the administration of the pilot interviews, during which the research team provided feedback to emphasize the following issues:

- ▶ Principles of interviewing, including sampling methods, question types, interviewing techniques, performance measurement, optimal interview times, and quality.
- ▶ Professional and ethical standards, including expected behavior of enumerators during the interview phone calls.
- ▶ Survey instrument review, including:

- ▷ Reasons why specific questions were included in the survey instrument and how they should be asked.
- ▷ Explanation of the terminology used throughout the questionnaire.
- ▷ Discussion on how to reply to difficult questions posed by respondents, including providing enumerators with a guide sheet with standard answers that they were to use.

4.2 Survey implementation: data entry, transmission, and delivery

The CATI application provided by the data collection company played a crucial role in the administration of the structured survey. The application included a comprehensive suite of survey scripts alongside a data entry interface to enable enumerators to conduct telephonic interviews with efficiency and precision. The application ensured a seamless flow for the SRHS survey process, with built-in mechanisms to prevent the omission of questions or essential instructions.

During data collection, respondents were asked whether they were willing to take the survey. If a respondent chose not to participate, the interviewer noted this, and the call ended. If a respondent was willing to take the survey, but requested to be called at a different time, the interviewer indicated this time in the dialer system, which reminded the interviewer to call back at the appropriate time.

Each segment of the script for each survey module was paired with corresponding data entry fields, allowing for the immediate recording of respondent replies. Upon the completion of each interview, the application automatically recorded the data, thereby streamlining the process for subsequent analysis and quality checks. The integrated dialer system further enhanced the process by automatically queuing the next contact, thus maintaining the momentum of the survey exercise until the target sample size was reached.

A dedicated management team monitored the administration of the survey, oversaw progress in real time to maintain data quality, and identified and rectified any discrepancies in the survey implementation process. This close supervision preserved the integrity of the survey data and facilitated the production of reliable and valid analytical outcomes.

Post-interview, the SRHS data was securely stored and formatted for consistency, ensuring a standardized database structure for straightforward data export. Personal identifiers were protected through the generation of unique IDs. The final dataset, inclusive of translated open-ended responses, was delivered to IFPRI in several statistical software formats.

4.3 Quality assurance and internal control mechanisms

Multilayered quality assurance measures ensured that the data collection process was transparent, reliable, and aligned with best research practices. The data collection process was designed to meet and uphold data integrity standards. In consequence, a series of automated and manual data quality checks were employed both during and after data collection. The following mechanisms were deployed:

- ▶ **Progress monitoring:** Analytic reports were extracted regularly from the CATI application. These detailed the time enumerators spent on individual questions and the

overall duration of their calls to respondents, as well as broader metrics like daily call volume and callback rates.

- ▶ **Automated data quality checks:** The system automatically verified incoming data against expected parameter ranges to ensure accuracy, immediately flagging to the enumerator any response that fell outside of the expected range. All responses were automatically verified for completeness.
- ▶ **Manual data quality reviews:** Post-interview, the SRHS data was scrutinized closely to identify any atypical response patterns or potential biases. The final dataset was subject to an exhaustive review. During this stage, the data was cleaned, coded, and analyzed for any anomalies or unexpected patterns that could indicate errors or biases.
- ▶ **Quality control calls:** A subset of respondents were called a second time by survey supervisors to validate some of the data originally collected by the enumerator.

For survey questions involving open-ended responses, enumerators received specific training on probing techniques to elicit more detailed and informative responses from participants. This training continued as needed through the survey administration period. Enumerators who consistently failed to meet data quality expectations were retrained or replaced.

These procedures ensured comprehensive data validation and facilitated the creation of a complete and accurate household survey dataset that was likely to provide reliable insights for informed decision-making for humanitarian response and rural development in Sudan under conflict conditions.

4.4 Privacy, data security, and ethics

Among the best practices in conducting survey-based research is establishing rigorous protocols to safeguard respondents' privacy and to ensure data security. Unique identifiers were assigned to survey participants to maintain anonymity and prevent any personal information from being linked to individual responses.

As part of the commitment of the research team to ethical research practices, the design and implementation of the SRHS conformed to the highest standards set by global research associations. Among others, these standards dictated that:

- ▶ Respondents for the SRHS had to be of legal consenting age—18 years or older.
- ▶ Respondents were fully informed about the survey's objectives and their rights. Based on this information, they then had to provide verbal consent to be interviewed.
- ▶ It was highlighted to all potential respondents that their participation in the survey was entirely voluntary. The respondents were informed of their right to withdraw at any stage of the survey administration process without penalty.
- ▶ SRHS enumerators were trained to understand and communicate the rights of participants, the confidentiality of their responses, and the measures that were being taken to protect data obtained from the participants.

To guarantee the security of the data collected, the SRHS data set is hosted on servers that comply with data protection regulations, including the European Union's data privacy laws. The servers are encrypted and housed within secure data centers to prevent unauthorized access.

Analysts' access to the dataset was segmented according to role-based permissions. All research team members undergo regular security awareness training. Procedures to be employed as a research team member included comprehensive background checks to ensure the reliability and discretion of any research team personnel handling the sensitive survey dataset. Additionally, systematic security assessments of the storage of the dataset are performed to regularly enhance defenses against potential cyber threats that would result in compromising the data and the privacy safeguards in place.

5. IMPLEMENTATION CHALLENGES

Translation and cultural adaptation: An initial challenge for the survey design pertained to the accuracy of the translation of the survey instrument. It was crucial that the questionnaire, originally drafted in English, be faithfully translated to accommodate local dialects and cultural nuances. The first translation attempt revealed discrepancies, prompting a comprehensive review. This was promptly addressed by leveraging the expertise of bilingual team members who conducted a meticulous retranslation, ensuring that the linguistic integrity and cultural relevance of the survey questions were preserved.

Data collection in remote areas: Reaching respondents in remote regions posed significant difficulties, particularly in areas with limited telecommunications infrastructure. This was compounded by the fact that many individuals had inactive phone numbers or faced network connectivity issues. To address this challenge, the data collection period was extended beyond the period originally planned, and call frequencies were increased. In addition, the survey team incorporated an additional set of contacts that had been used in a recently completed agricultural survey, thereby expanding the pool of potential respondents.

Respondent displacement: Displacement of residents from Khartoum due to conflict was a significant challenge to completing interviews. This led to initial difficulties in meeting the sample quotas. Over time, and due to the supplementary contact list, which was integrated into the survey process, the response rate improved, and quotas were successfully met.

Incentivizing participation: Initially, the survey team encountered challenges in providing incentives to participants due to logistical constraints. Recognizing the importance of incentives in motivating participation, all respondents received a gratuity of additional mobile telephone airtime.

Network instability in Sudan: There were instances of network instability during the survey administration period. This impacted the response rates in certain states and also impacted real time syncing of interviews. As a result of the syncing problem, the survey enumerators slightly exceeded the target sample size in several states. These surplus cases were dropped from the final dataset during the data-cleaning process.

ANNEXES

Annex 1—Household survey sample design factors

Purpose of the survey: The sample of the SRHS, which is designed as a panel survey, is expected to track selected outcome indicators over time. These indicators include agricultural productivity, poverty, food security, and nutritional outcomes. The level of inference, i.e., reporting level, is at the state-level.

Cluster size: Sudan has 18 states and 189 localities. Enumeration areas, which are based on sub-localities, would be a natural primary sampling unit for the SRHS. However, given that a telephonic CATI method was used to implement the SRHS, the larger localities were used as the sampling units.

Primary indicator: The size of the sample was in part determined by a set of primary indicators for the survey. Ideally, a target indicator with the highest intra-cluster correlations from a relatively recent household survey is used in a power calculation to compute the sample size required to produce credible estimates for all indicators of interest. Poverty incidence from the 2014 MICS was the target indicator used for the power calculation to determine the appropriate sample size for the SRHS.

Minimum detectable effect size: Sample size depends on the minimum level of change (known as minimum detectable effect size) that is to be detected in the relevant indicator at a specific level of statistical significance. Smaller effect sizes require larger survey samples.

Significance and power: The sample must be sufficiently large to minimize the likelihood of detecting an effect that does not exist (statistical significance) and to maximize the likelihood of detecting an effect that does exist (statistical power). Following standard practice, for the SRHS, these were set at a level of significance of 5 percent (two-tailed or one-tailed) and statistical power of 80 percent.

Design effect: The design effect (Deff) reflects the extent to which the indicator of choice is correlated across households or individuals within a specified group or cluster, usually defined by geographic location. Intra-cluster correlations and cluster-level sample size are used to measure this extent. Higher intra-cluster correlations mean that the design effect is stronger and that larger samples will be needed for the survey. Given clustering, it is necessary to have enough clusters to obtain robust tests. Towards that end, the intra-cluster correlations of possible outcome indicators were considered first. The conservative compromise was to use the indicator with the highest intra-cluster correlations in the power calculations.

Attrition: Over time in a panel survey program, some sample households will move to other localities, others will dissolve with members dispersing into existing or new households, and others may choose not to continue to be interviewed and drop out of the survey program. Given the current context of conflict in Sudan, we expect attrition in subsequent rounds of the SRHS to be relatively high. However, estimating attrition will only be possible once the second round is completed.

Annex 2—Sample size determination for the 2023 Sudan Rural Household Survey

Two relatively recent and relevant sources of household information could have been used in computing the sample size for the 2023 SRHS—the 2022 Sudan Labor Market Panel Survey (SLMPS) and the 2014 Sudan Multiple Indicator Cluster Survey (MICS). Both surveys offer insights into poverty, welfare, food security, nutrition, and livelihoods, providing state-level estimates of these outcomes. The MICS also provides some relevant parameters for determining an appropriate sample size for the SRHS, including design effects and attrition rates.

In the end, data from the 2014 MICS was used in the sample size calculations for the 2023 SRHS. Unfortunately, data from SLMPS 2022, the more recent survey, was not available for the IFPRI research team to use until after the sample design analyses had been completed.

In computing the necessary sample size, the objective was to identify a sample large enough to allow a comparison of proportions or means based on estimates of indicators drawn from the analysis of data from the 2014 MICS. A set of additional assumptions were applied to the analysis, including an estimate of design effect from MICS and consideration of possible intensification in the detrimental consequences of the multiple shocks affecting Sudanese society since 2015—political crises, COVID-19, and the international commodity price effects of the war in Ukraine. A budget constraint was also part of the calculation. The desired inference level for the SRHS is the state.

The key indicators from the 2014 MICS used in the sample size analysis for the 2023 SRHS were the poverty headcount ratio—the incidence of poverty based on household consumption expenditures—and the prevalence of stunting among children under five years of age.

The poverty analysis of the 2014 MICS estimated a poverty headcount of 28.3 percent for the four states of Khartoum, Gezira, Blue Nile, and South Kordofan and an under-five stunting prevalence of 34.5 percent. However, these rates are judged to have likely increased since 2014 due to the repeated shocks Sudan has experienced. Second, several experts on poverty and malnutrition in Sudan consulted during the SRHS design phase suggested that these levels were underestimated even for 2014. Third, more recent poverty headcount and child stunting estimates show that these rates are likely on an upward trend.³ Given these considerations, we assumed a national poverty headcount rate of 50 percent and a national under-five stunting rate of 50 percent in 2022.

The desired minimum detectable effect size (MDE or δ) for the SRHS was an 8 percentage point change in poverty headcount and a 9.5 percentage point change in child stunting over three to five years, equivalent to about a 16 percent and a 19 percent change in poverty and child stunting, respectively, over the period. The statistical significance value (α) used for the sample size computation is 5 percent for a two-tailed test, i.e., $z_{1-\alpha}$ of 1.96, where z is the cumulative normal variate. A power value ($1-\beta$) of 80 percent was used, corresponding to a $z_{1-\beta}$ of 0.84. A non-response or attrition (r) rate of 7.5 percent was assumed, with this value based on the experience with the 2014 MICS. The same survey computed a design effect

³ Dabanga Sudan. 2022. *Sudan's poverty rates potentially even higher than reported*. July 1, 2022. <https://www.dabangasudan.org/en/all-news/article/sudan-s-poverty-rates-potentially-even-higher-than-reported> (Accessed March 6, 2023).

(Deff) of 1.7, and the same value was assumed for the sample size computation for the 2023 SRHS.

The sample size required by the survey (n_0) is calculated using the following formula for proportion or ratio type variables:

$$n_0 = \text{Deff} * \left\{ \sqrt{\frac{z_{1-\alpha} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)}}{\delta}} \right\}^2$$

$$\bar{p} = \frac{p_1 + p_2}{2}$$

where p is the estimated level of the proportion of interest (poverty or stunting), with subscripts 1 and 2 representing initial (baseline) and final (end-line) levels, respectively. The rest of the variables in the formula are as defined earlier in this annex.

Two adjustments need to be made to the initial sample size computed. First, the household sample must allow a sufficient number of relevant persons to be contacted in the case of individual-level indicators. This applies to the key variable of stunting in the present context and reflects the share of under-five children in the household and the average household size, both values of which are obtained from the 2014 MICS database. Second, the sample size needs to accommodate expected household non-response and attrition.

Once the political crisis in Sudan is resolved, it is reasonable to expect that the economy and the well-being of Sudanese households will improve rapidly. Accordingly, relatively large MDEs were assumed, with the MDE for the stunting indicator being larger. Nevertheless, the sample sizes for the SRHS computed using the two key variables, poverty headcount and child stunting, differed by only three households. Based on these calculations, the planned SRHS sample was rounded to 4,500 households nationally.

Finally, It is important to note that the overall population size is irrelevant to the size of the sample required. In other words, regardless of the varying sizes of two or more populations, a single sample size can accurately represent them all, provided that the underlying determinants of the sample size remain consistent across the populations.

ABOUT THE AUTHORS

Oliver K. Kirui is a Research Fellow in IFPRI's Development Strategy Governance (DSG) Unit, based in Khartoum, Sudan. **Mosab Ahmed** is a Senior Research Assistant in IFPRI's DSG Unit, based in Khartoum. **Alemayehu Seyoum Taffesse** is a Senior Research Fellow and Ethiopia Strategy Support Program Leader in IFPRI's DSG Unit, based in Addis Ababa, Ethiopia. **Hala Abushama** is a Research Assistant in IFPRI's DSG Unit, based in Khartoum. **Khalid Siddig** is a Senior Research Fellow and Sudan Strategy Support Program Leader in IFPRI's DSG Unit, based in Khartoum and an Associate Professor at the University of Khartoum, Sudan.

ACKNOWLEDGMENTS

This study was prepared with the financial support of the United States Agency for International Development (USAID). The authors are grateful to colleagues from UNDP—Vito Intini, Walid Merouani, Fekadu Terefe, Nada Nohra, Abdalatif Hassan, Shireen AlAzzawi, and Karin Ruf—and WFP—Siddharth Krishnaswamy, Cinzia Monette, Alberto Gualtieri, and Alice Clough—for their review of the survey instrument and constructive suggestions throughout the survey design and implementation process. The authors also thank the teams at the survey data collection firms contracted to assist in the implementation of the 2023 SRHS—at GeoPoll, Purity Njiru, Zipporah Moraa, John Murunga, and Christine Delmeiren, and at Consultsat, Mohamed Hamza and Doaa Maysara. These colleagues provided critical technical details and useful suggestions on the sampling procedure and implementation of the survey.

The Sudan Strategy Support Program (SSSP) is managed by the International Food Policy Research Institute (IFPRI) and is financially supported by the United States Agency for International Development (USAID). This publication has been prepared as an output of the SSSP and has not been independently peer-reviewed. Any opinions expressed here belong to the author(s) and are not necessarily representative of or endorsed by IFPRI.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

A world free of hunger and malnutrition

IFPRI is a CGIAR Research Center

IFPRI Sudan, | Email: ifpri-sudan@cgiar.org | <https://sudan.ifpri.info/>

© 2024, copyright remains with the author(s). All rights reserved.