2 Methodology

Bymolt, R., Laven, A., Tyszler, M. (2018). Demystifying the cocoa sector in Ghana and Côte d'Ivoire. Chapter 2, Methodology. The Royal Tropical Institute (KIT).

2.1 Research approach

The research methodology was designed by KIT with input from our funding partners. The over-arching aim of the study is to contribute to the cocoa sector's body of knowledge and provide a solid evidence base to test common assumptions and beliefs.

The research covers a broad range of subjects including household demographics, food security and nutrition,[1] and crop choices and crop diversification. Specific to cocoa, the study investigates why households grow cocoa, and analyses the major aspects of cocoa production and marketing. Household wealth, income and poverty is also assessed, and further disaggregated in a cluster analysis. The study also looks at intra-household dynamics which are brought together in a gender chapter.

Originally, we aimed to integrate three elements in particular. These are 'on farm', 'between farm' and 'within household'.

On farm: This element involved analysing all aspects of cocoa production and marketing. We also sought to understand the relative importance of cocoa in relation to other 'competing crops', and the role of crop diversification within the farming system.

Between farm: This element involved analysing household demographics, and disaggregating households in the sample according certain characteristics. (For example, cocoa vs non-cocoa households; male headed vs female headed households, etc.)

Within household: This involved understanding intra-household dynamics, particularly male and female roles on the farm and in the household, as well as access to assets and decision making power.

An important consideration was to make the cocoa household (rather than the individual farmer) our unit of analysis. This is because we recognise that household members often take on different roles in cocoa production and marketing, as well as different roles in other income-generating activities, all of which contribute to household wellbeing.

[1] The Sustainable Trade Initiative (IDH) and the Global Alliance for Improved Nutrition (GAIN) have signed a Memorandum of Understanding, to mark a new partnership to improve the health and nutrition of smallholder farmers and their communities in agricultural value chains. Our research findings provide input to the development of their joint strategy and company programmes in Ghana and Côte d'Ivoire.

A mixed-methods approach was chosen for this research involving 1500 household surveys in Ghana and 1500 in Côte d'Ivoire, as well as 37 focus group discussions in each country.

For the household survey, a large sample size was deemed important so that data could be analysed for statistical significance, and findings could be potentially be generalised at the regional and country level. The household survey covered demographic information, socio-economic characteristics, nutrition and food security questions and detailed questions on the production and sales of each household's two most important crops. The survey borrowed from other indexes where relevant, including the Dietary Diversity Score (DDS)[2] and the Poverty Probability Index (PPI)[3] and included DHS Wealth Index survey questions.[4] The household survey was written in XLSform markdown and deployed on digital tablets running Open Data Kit. The survey has been made available for download on the project website, along with the database.

Focus group discussions were also conducted in the same communities as the household surveys. Discussions focussed on the understanding the reasons why households make particular decisions on the farm and in the household, and what the current trends are in different communities. Focus group discussions consisted of a variety of different exercises that complemented the household survey questions. The participatory exercises included scoring and ranking, and provided the opportunity to probe farmers' perceptions to understand risks and their behaviour.

Data was collected in Ghana between November 2016 and January 2017, and in Côte d'Ivoire between February and March 2017. In the household survey, respondents were asked to recall information from the 2015-2016 cocoa season. In Ghana, this timing coincided with elections in December 2016, which resulted in changes in the government and within COCOBOD. In early 2017, there was a major price-drop in the world-market price of cocoa. At the time of fieldwork, this was yet to directly affect farmers in our sample.

[2] FAO and FHI 360 (2016). Minimum Dietary Diversity for Women: A Guide for Measurement. Rome: FAO. Available at http://www.fao.org/3/a-i5486e.pdf

[3] Portal on Poverty Probability Index by Innovations for Poverty Action (IPA). Link https://www.povertyindex.org/about-ppi

[4] Portal on the Demographics and Health Survey Programme. USAID. Link https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm

2.2 Research phases

The research was conducted in three phases:

- Desk study and methodology development
- Fieldwork
- Analysis and reporting

2.2.1 Phase 1 - desk research and methodological development

An extensive and systematic desk study was conducted to understand the present state of knowledge of the cocoa sector in Ghana and Côte d'Ivoire. The desk research phase also allowed the researchers to identify areas where there were inconsistencies across studies, or where certain beliefs did not seem to be well supported by strong evidence.

Approximately 100 key articles were collected and coded in ATLAS.ti. A summary report was then made structured around the main coded topics. The research team are grateful to our research and funding partners for their contributions and for sharing a number of key articles with us.

2.2.1.1 Research questions

The desk research served as a basis for developing an initial list of hypotheses and assumptions to be tested. This, in turn, led to the development of overarching research questions:

- What are the defining demographic and socio-economic characteristics of cocoa and non-cocoa producing households?
- What are the dominant crop/livelihood options in the research areas, and why?
- Are cocoa households leaving cocoa or increasing the share of land under cocoa, and how is the process happening?
- To what extent are cocoa household incomes diversified, and in what ways?
- What is the poverty and wealth status of cocoa households compared with non-cocoa households?
- What is the nutrition and food security status of cocoa households compared with non-cocoa households? What is the availability, affordability of different food groups?
- To what extent do households invest in inputs and apply good agricultural practices for cocoa and other crops?
- What are the costs, revenues and profitability of cocoa compared with other crops?
- How do cocoa households sell the cocoa, and how are cocoa prices formed?
- What roles do men and women typically play on the farm and in the household, and why?

- Who typically makes decisions about investments in cocoa and expenditures on other household items?
- How are cocoa institutions perceived and what can be done to better support sustainable production?

2.2.1.2 Development of research tools

The household survey and focus group discussion exercises were developed following the desk research, cross checked with our research questions and hypotheses, and were validated with our research partners.

An extensive household survey was developed to cover each household's social-economic characteristics, nutrition and food security status, agricultural and non-agricultural income sources, and the production and sales of their two most important crops. The survey was also designed to capture information on savings and loans, household assets, livestock assets and productive assets. The survey was designed to be around one hour long to prevent respondent fatigue.

The survey tool was developed in XLSForm markdown language and deployed on digital tablets running Open Data Kit (ODK) software.[5] ODK was chosen for its ease of use and accuracy of data collection and aggregation compared with traditional paper-based surveys. For example, the survey was programmed to make live calculations, which then gave enumerators error and warning alerts when unexpected values were entered. Programmed skip logic was used so that respondents were only asked questions consistent with their prior responses, whilst non-relevant questions were automatically skipped. Furthermore, there is no need to transcribe data from paper to digital format, saving time and eliminating the potential for transcription errors. Once completed and checked, surveys were pushed to a cloud server to provide remote backups and monitoring of progress.

To enhance the quality of survey data, we often prefaced blocks of questions with a 'do you know' question. For example, respondents were asked *'Do you know how much land your household used to cultivate all your crops last year?'* Binary yes/no questions were used in conjunction with programmed skip logic on the tablets. If respondents answered 'yes' then they would receive the block of questions. If they answered 'no', then the block would be skipped, preventing respondents from giving insufficiently informed responses on the topic. Data quality was also enhanced by allowing respondents to answer certain questions in any unit they felt most comfortable. For example, land sizes could be answered in acres, hectares or poles, and volumes could be answered in bags or kilograms.

[5] See Open Data Kit: https://opendatakit.org

Focus group discussion exercises were developed to complement the household survey. Whilst the household survey tried to gather data on 'what' and 'how much', the focus group exercises were designed to answer the questions 'why', 'how', and 'for whom'. The focus group exercises were designed to be participatory to provide participants with the opportunity to share their perceptions in detail. Many included ranking and scoring components. The exercises were methodologically underpinned by the PADev method.[6] Men and women both attended the focus group discussions but sat apart. The facilitator took turns prompting men and women to respond during each discussion. Each focus group session took around 4-5 hours, and included the same participants who took part individually in the household survey.[7] Four focus group exercises were conducted in each community, due to time availability. The researchers rotated the exercises between communities. The following exercises were included in the study:

- **Income sources:** identifying, ranking and describing the most important income sources in the community;
- **Calendar:** understanding income and expenditure patterns throughout the year;
- **Nutrition:** discussing availability and affordability of food groups, and understanding household decision-making roles;
- **Statements:** Likert scale scoring and discussion on a selection of research hypotheses;
- **Budget:** Participatory budget exercise focusing on labour days, input costs, yield and price as a way to triangulate survey data and generate initial estimations of profitability;
- **Changes:** understanding participant perceptions of changes and trends in a range of domains from five years ago to today;
- **Institutions:** discussion of institutions present in the community and scoring relative satisfaction/dissatisfaction with the products or services they provide;
- **Intra-household relations:** understanding household roles and input in productive decision- making (based on the A-WEAI);[8]
- **Access:** understanding ownership of assets, access to and decisions on credit, control over use of income and group membership (based on the A-WEAI).

2.2.1.3 Sampling

A sampling frame was developed during phase 1. The desire for the research to be representative and generalisable necessitated an approach to sampling that was both random and practically implementable in the field with the resources available.

Geography was the main stratification characteristic. The geographic focus in both Ghana and Côte d'Ivoire was on cocoa growing areas (rather than cocoa households

[6] See www.padev.nl

[7] Participants were provided with drinks and snacks, and later a lunch. Normally the 4-5 hour session would start early in the morning (8 am-1 pm).

[8] The WEAI is a survey-based index designed to measure the empowerment, agency, and inclusion of women in the agricultural sector. Source: Malapit,H. Kovarik, C., Sproule, K., Meinzen-Dick, R. and Quisumbing, A. (2015). Instructional Guide on the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI). Available at https://www.ifpri.org/sites/default/files/a-weai_instructional_guide_final.pdf

specifically). The intention was to sample 1,500 respondents in both Ghana and Côte d'Ivoire (total 3,000). This was to be achieved by visiting one community per day, and carrying out approximately 40 surveys in each community.

Randomised sampling was done in two stages. First, a long list of cocoa growing regions was compiled for both countries. In Ghana we gave more weight to the Western Region, because it is a region with a high proportion of Ghana's total cocoa production. In the 2015/2016 season, the Western region produced 53% of Ghana's total production.[9] Computer-generated random numbers were assigned to each community in each district of each region. In both Ghana and Côte d'Ivoire, the community randomly assigned the highest number in each district was selected for the study. This ensured that we did not simply visit communities that were the most accessible or convenient.

At the community level, randomisation was carried out by one of the research team members a few days in advance of each visit. The team member worked with local leaders in the communities to select a random sample of people in each community. This was done through a simple transect walk, north, east, south and west from the centre of the village. Ten households from each direction were selected, skipping every *Nth* house, depending on the size of the community. This ensured that potential bias was removed to avoid the systematic selection of the best farmers, those who are well-connected members of groups, or those who frequently participate in projects.

It is important to note that there was no requirement that the household had to include a cocoa farmer, or even any kind of farmer. We did not intentionally select respondents that were involved in a certain programme or affiliated with a certain buyer, as is often the case in evaluations of sustainability programmes.

The only requirement was that no less than one third of respondents was female, which was important both for having women's voices in the research and, specifically, for measuring women's dietary diversity.[10] In the final sample, in both countries, 34% of respondents were women.

In total, 1,560 household surveys were collected in Ghana and 1,485 in Côte d'Ivoire (total 3,045 household surveys). In each country, 37 focus group discussions were conducted, one in each of the communities visited (total 74 focus group discussions). The same sample of participants attending the focus group discussions were interviewed for the household survey.

[9] COCOBOD (2018). Regional Cocoa Purchases. Available at https://cocobod.gh/weakly_purchase.php

[10] FAO and FHI 360 (2016). Minimum Dietary Diversity for Women: A Guide for Measurement. Rome: FAO. Available at http://www.fao.org/3/a-i5486e.pdf

Figure 2.1 Location of sampled communities

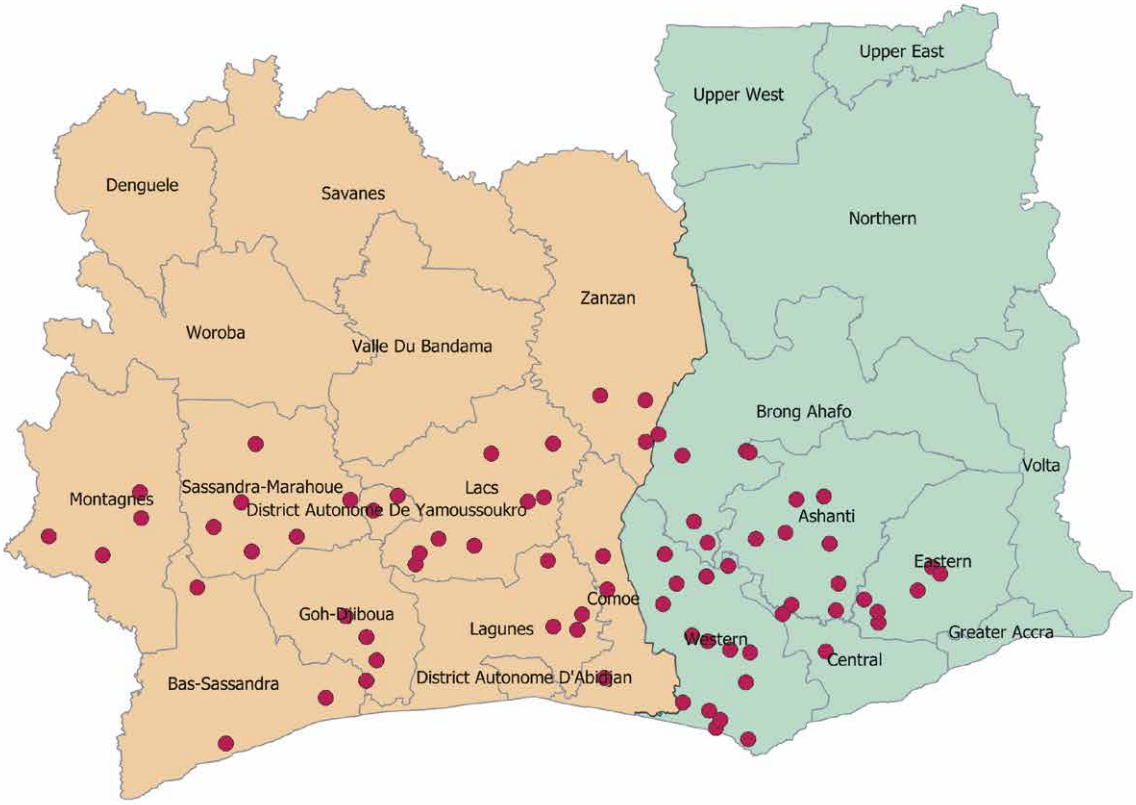

Table 2.1 Regions sampled in Ghana, percent of respondents

Region	% respondents	N respondents
Ashanti	20%	317
Central	5%	72
Brong Ahafo	17%	270
Western	41%	645
Eastern	16%	256
N		1560

Note: percentages may not add up to 100% due to rounding

Table 2.2 Administrative Districts[11] sampled in Côte d'Ivoire, percent of respondents

Region	% respondents	N respondents
Autonome De Yamoussoukro	6%	90
Lacs	24%	358
Montagnes	10%	142
Bas-Sassandra	11%	159
Goh-Djiboua	9%	136
Zanzan	8%	126
Sassandra-Marahoue	17%	251
Comoe	7%	97
Lagunes	8%	126
N		1485

Note: percentages may not add up to 100% due to rounding

[11] An administrative district in Côte d'Ivoire is approximately the same unit as a region in Ghana

2.2.2 Phase 2 - Fieldwork and data collection

In each country, local enumerators, managed by a local consultant, conducted the household surveys with respondents. Prior to fieldwork, KIT delivered three days of intensive training to enumerators in each country. The goal of the training was to familiarise the research team with the research project, train enumerators on the use of tablets and to ensure that the meaning of all questions was understood in detail. Training was followed by a testing phase in a community to address any remaining issues.

The research team arrived at each research location early in the morning and liaised with the village leader, who had been informed of our visit in advance, and who had supported the random sampling process. Following introductions with the community, the lead KIT researchers began the focus group discussion exercises.

Facilitators took particular care to involve female participants full in the focus group discussions. For example, women were often asked to give their perspectives before the men. Care was also take to balance the involvement of 'opinion leaders', who are often well-respected and knowledgeable members of the community, with the voices of 'ordinary' farmers. A plurality of opinions was welcomed in the focus group discussions.

During the focus group discussions, the enumerators would each take one person out of the discussion to do a household survey with them individually. On completion the respondent would return to the focus group discussion and the next person would be surveyed. For most household surveys, female enumerators surveyed female respondents.

The survey and focus group discussions were always conducted in a location chosen by the community. In Ghana, the research was often conducted in a local church or community centre. In Côte d'Ivoire, once the chief has granted permission, the participants were gathered in a central, shaded place (sometimes under the central mango tree).

For the first 10 days, every completed household survey was checked by a KIT advisor for any errors or misunderstandings. Daily briefings were held at the beginning of each day to discuss with the team any problems that may have occurred the previous day. Thereafter, when enumerator errors had fallen to very low levels, a sample of each

2.2.3 Phase 3 - Analysis and reporting

The report has been divided into the following chapters. Following the introduction (Chapter 1) and methodology (Chapter 2), we first describe the characteristics of respondents and households in our sample, giving attention to demographics such as age, gender of the household head, education levels and so on (Chapter 3). In

various places in the report we disaggregate by these characteristics as we analyse differences between groups. In Chapter 4, we analyse the nutrition and food security status of households. In Chapter 5 we present an analysis of crop choices and crop combinations, followed by a detailed chapter on land (Chapter 6). Chapter 7 presents a qualitative description of 'why households produce cocoa'. Chapter 8 presents data on cocoa production activities, labour and inputs. In Chapter 9, we continue with data on cocoa farmer groups, certification and training. Productivity and yields are analysed in Chapter 10. Chapter 11 gives a detailed description and analysis of cocoa production and marketing. Other major crops are compared with cocoa in Chapter 12. In Chapter 13 we analyse wealth, poverty and income. Chapter 14 summarises gender issues presented throughout the paper and provides some additional analysis of intra-household relations. A cluster analysis is presented in Chapter 15 in which we address the question of whether or households can be distinguished in certain categories. The main findings and conclusions of the study are brought together in Chapter 16.

Throughout the report, we have always disaggregated Ghana and Côte d'Ivoire data. In addition we have presented a number of tables disaggregated in other ways, such as by the sex of the household head, or by cocoa or non-cocoa households. Other disaggregation is presented where statistically significant differences were found.

The format we have chosen for most tables presents the coefficient for the two countries, the standard error (of the mean), the P-value and an indication of its statistical significance using stars. ** means statistical significance at the 5% level (the most commonly used indication that a difference in means between two groups is not attributable to chance), and *** indicates very high significance at the 1% level. At the end of most tables, we also indicate the number of observations (N). Data with low N have often been excluded because an interpretation is usually not robust. Finally, we usually include the variable name at the bottom of the table so that others can replicate the findings if they download the database.

When we present an association between two categorical variables, we have used a Chi-squared test (for example, to examine categories of educational level achieved by the sex of the household head). When analysing the association between a continuous variable and categorical variables, we have used a one-way ANOVA test (for example, to analyse land size by regions of a country). Similarly, if the categorical variable has only two groups, the one-way ANOVA is essentially the same as an unpaired t-test analysing the difference of means (for example land size by the sex of the household head).

The full dataset from our research is published on Dataverse[12] for other researchers to freely download and use for their own research. To facilitate the use of Dataverse

[12] https://doi.org/10.7910/DVN/82TWZJ

we have included variable names under each table presented in this report. We also provide other readers with a 'data navigator' which allows any user to browse basic descriptive data in their web browser.

2.2.4 Limitations

In the set-up of our research we have tried to be as extensive as possible, however, like all studies there were still limitations. In some instances, this may suggest the need for more focussed research to answer highly specific questions.

We have given detailed attention to cocoa in this report, as it is the main crop of interest. We also collected detailed data on eight other major crop options in cocoa producing areas. However, during analysis we found that for several crops, the number of observations too low to confidently provide robust details of the production system, or revenue and profit. In other cases, respondents found it particularly difficult to recall their total production. A case in point was cassava, which is typically not harvested at once but bit by bit. Some respondents even sell the field and leave it to the buyer to harvest, which makes estimates even more complicated.

For disaggregated data, we also sometimes encountered problems with a low number of observations. For example, there is a low number of female-headed households producing cocoa in Côte d'Ivoire. In some cases, this makes it difficult to make definitive statements about certain groups or sub-groups and, in other cases, we have relied on the qualitative focus group data to describe the findings.

Finally, the research did not cover some important topics, such as child labour. This would require a specific methodological focus to do justice to the issue. We would have liked to have been able to collect more detailed survey questions on aspects like food security and household expenditures, however we judged that this would make the survey too long for respondents.